# INSIGHTS INTO EARLY DEVELOPMENT FROM GENETIC AND EPIGENETIC DATA ACROSS TISSUE TYPES

By

Shan V. Andrews

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

April, 2018

# ABSTRACT

The etiology of autism spectrum disorder (ASD) remains largely undiscovered. Common genetic variants are increasingly being realized, but the functional biology implicated by these SNPs has yet to be elucidated. Integrating epigenetic data, particularly DNA methylation (DNAm), with ASD SNP data is warranted, given its success in schizophrenia and bipolar disorder and the evidence for an epigenetic basis to ASD. Given that DNAm is tissue-specific, it is of interest to consider DNAm from brain, the main tissue of interest for ASD, and peripheral tissues simultaneously so that SNP and CpG-based insights can be compared across tissues. One peripheral tissue, the placenta, is understudied in an ASD context, despite being the critical organ regulating the gestational period, a key window of interest for ASD etiology. Given this role, the placenta also "programs" early and later life phenotypes, and sex-biased ratios in many phenotypes, including ASD, may be driven by sex-specific placenta functions. In **Chapter 1**, we first describe the current literature on ASD etiology, and then review placenta structure and function, role in fetal brain development, and known differences by fetal sex. Next we use DNAm data across tissues in conjunction with paired genotype data to find that, across blood and brain, ASD-associated SNPs are enriched for methylation quantitative trait loci (meQTLs), or SNPs that are associated with DNAm, and that the DNAm sites controlled by ASD-related SNPs are engaged in immune response pathways (**Chapter 2**). We then focus on studies of the placenta, first demonstrating the additional enrichment of ASD-associated SNPs with placenta meQTLs (**Chapter 3**), and conducting queries for differentially methylated regions (DMRs) and global methylation differences related to ASD (**Chapter 4**). Finally, we discover a genome-wide significant DMR for fetal sex in placenta near *ZNF300*. DNAm levels in this region are related to placenta morphology, demonstrating a mechanism by which sex-specific placenta structure, and subsequent sex-biased vulnerability to disease, may occur (**Chapter 5**). This dissertation

contributes to our understanding of how epigenetics mechanisms, particularly acting in conjunction with genetic variation, contributes to understanding of ASD etiology across tissues, and govern placental functions to create sex-specific gestational environments.

Thesis Readers:

Christine Ladd-Acosta, PhD (Advisor)

M. Daniele Fallin, PhD (Advisor)

Dan Arking, PhD

Debashree Ray, PhD

Alternates: Alden Gross, PhD and Xiaobin Wang, ScD

# ACKNOWLEDGEMENTS

First I would like to thank my advisors, Chris Ladd-Acosta and Dani Fallin. I am extremely grateful for your mentorship for the past 6 years during my master's degree and then of course my PhD. Thank you for always being open to new ideas, no matter at what stage in development you heard them, and for teaching me how to blend epidemiology, genomics, and bioinformatics in the right ways to lead to meaningful work. I know I am the researcher I am today due to both of your guidance. I am proud of the work we have accomplished together and I will miss working with you both.

Next I would like to thank Kasper Hansen, who has been a great additional mentor to me. I always appreciated your frank criticism and I always felt that I walked out of your office with a very tangible improvement in my ability as a scientist. I would like to also thank you specifically for improving my ability to write R code; I greatly appreciated your insights into how to best write concise, user-friendly functions and I know you have given my knowledge that I will carry with me for years to come.

I have additional mentors I would like to thank as well. First, to Dan Arking, thank you for all of your advice over the years. I always marveled at your ability to be able to give great insights and helpful next steps after looking at my results for only a few minutes. Next, to Andy Feinberg, thank you for all your insights, I always felt very lucky to gain your perspective on my work. To Kelly Bakulski, thank you for your friendship, support and guidance over the years. It is often said in academia that it is important to have someone at the stage ahead of you who you can look up to and who can serve as a role model, and you were undoubtedly that person for me. Finally I would like to thank Kelly Benke and Brion Maher, who I started my first research project with in

the fall of 2012. Thank you setting me off on the right foot and easing me in to the research process.

To Shannon Ellis – thank you being a wonderful collaborator and friend. I learned so much in our meetings together figuring out how to do our various meQTL analyses and I will always be very grateful for that time we spent working together.

I need to give a special thanks to all the study staff and participants for the Study to Explore Early Development and the Early Autism Risk Longitudinal Investigation. In particular I would like to thank Jamie Kaczaniuk and Michelle Landrum, who always made navigating study requirements and logistics very easy and who are some of the friendliest faces at the School of Public Health.

Thank you to the Maryland Genetics, Epidemiology, and Medicine Training program and all of the faculty behind it, in particular Priya Duggal, Dave Valle, Dani Fallin, and Jennifer Deal. I have gained numerous experiences I would not have had otherwise through this program and I know I am a more well-rounded researcher because of it.

Thank you to the JHSPH Epidemiology department, in particular the genetic epidemiology faculty who have taught me and who I have TA'd for, including Priya Duggal, Terri Beaty, Robert Wojciechowski, and Linda Kao. To Ali Abraham and Bryan Lau, thank you for giving me the opportunity to TA for you for several years. I learned so much about the best ways to teach students and I always enjoyed your company.

Thank you to all of my friends for their support during this process, in particular John, Robby, and Chet, and all of my roommates during my time in Baltimore. Many of you could relate to my

experiences first hand, and even if you didn't, I always felt privileged to have you in my life as an outlet during the more difficult times.

Thank you to my sisters for your love and support throughout my life and especially in this PhD process. You two have always been wonderful role models for me. I respect you both so much and am proud of everything you have accomplished.

I must say a very deep thank you to my parents, to whom I don't say that nearly enough. Thank you for immigrating to this country and giving me opportunities and experiences I would not have had otherwise. Thank you for all of your hard work during my childhood and for fostering a supporting environment. I am grateful for all you have given me and I hope to now start repaying the favor.

And finally thank you to my dearest Brooke, my future wife and my best friend. I can't describe what a joy it is to have you in my life and how lucky I was throughout this process to know that no matter what happened during the day, I could always come home to you. Thank you for all of your support and, of course, for bringing Leonard into my life. As I've always said, I'm not sure if I'm a dog person now, but I'm definitely a Leonard person.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

## 1.1 Autism spectrum disorder

### *1.1.1 Definition*

Autism spectrum disorder (ASD) is characterized by social, behavioral, and communication deficits and, as of 2012, it affects 1 in 68 children in the US. Estimated prevalence is higher in males (23.6 per 1,000 individuals) than females (5.3 per 1,000 individuals) (Christensen et al., 2016). There is also a high degree of phenotypic heterogeneity in ASD, underscored by various comorbid conditions relating to intellectual development, language ability, and attention deficit hyperactivity disorder, among other traits (Gillberg & Fernell, 2014).

Prevalence for ASD has increased rapidly worldwide in recent decades, though the most recent estimate provided by the Autism and Developmental Disabilities Monitoring Network remained the same from the estimate of two years prior (Christensen et al., 2016). A recent Danish-registry based study attempted to find the attributable fraction of this prevalence increase to improved diagnosis and reporting practices vs. etiologic factors. This study concluded that the majority (60%) of the increase in reported ASD prevalence can be attributed to both changing in diagnostic criteria and the inclusion of outpatient controls (Hansen, Schendel, & Parner, 2015). Thus, a substantial fraction of the ASD prevalence increase in not explained by secular and diagnostic trends and likely represents a true increase in the disease prevalence. Studies are needed to understand the etiology of ASD and identify risk factors that can be intervened on to reduce the burden of disease.

### *1.1.2 ASD heritability estimates*

Early estimates of ASD from twin or family-based studies have varied greatly, ranging from 38% to 95% (Colvert et al., 2015; Hallmayer et al., 2011; Tick, Bolton, Happé, Rutter, &

Rijsdijk, 2016). A recent Swedish registry-based analysis sought to determine the relative contribution of common and rare genetic variation to overall liability. They estimated the SNP-based heritability, or that due to common variation at roughly 49%, and a total genetic contribution (i.e. including rare inherited and *de novo* variants, and non-additive genetic factors) of about 60% (Gaugler et al., 2014), thus, leaving approximately 40% of ASD liability being attributable to environmental factors.

### *1.1.3 Environmental risk factors associated with ASD*

Several demographic factors have been linked to ASD. Strong evidence indicates that advanced maternal and paternal age is a risk factor for ASD (Wu et al., 2017). Also, it has been consistently reported that there is a higher prevalence of ASD in males than females, and that males are particularly over-represented in higher-functioning cases. Male-female ratios of ASD tend to equilibrate when limiting to individuals who have comorbid intellectual disability. Overall, research has increasingly indicated that ascertainment biases may exaggerate the frequently stated 4:1 male to female ratio, but that true sexual dimorphism in ASD exists biologically (Halladay et al., 2015; Werling & Geschwind, 2013). It has been hypothesized that there exists a "female protective effect" in ASD such that females tend to accumulate a higher number of ASD risk factors before reaching a diagnostic threshold (Robinson, Lichtenstein, Anckarsäter, Happé, & Ronald, 2013). Genetic studies showing a higher burden of ASD-related genetic variants in females support this notion (Jacquemont et al., 2014). However, still the mechanism of sex differences in ASD is not completely understood. In **Chapter 5**, we query for epigenetic signatures related to sex in placenta, in an effort to understand sex differences in the gestational period that can program sex-biased phenotypes such as ASD.

There has also been more recent focus on the contribution of environmental factors to ASD etiology after recent, more rigorous, studies have shown the initial very high (~90%) heritability estimates were overestimates, thus supporting the notion of a stronger environmental risk component than initially thought (Hallmayer et al., 2011; Sandin et al., 2014). Evidence supports the prenatal period as a critical window of environmental exposure risk for autism (Lyall, Schmidt, & Hertz-Picciotto, 2014). Therefore, many maternal lifestyle factors have been examined for their potential role in contributing to ASD. Reduced interpregnancy interval has been shown to be associated with ASD (Cheslack-Postava, Liu, & Bearman, 2011). Prenatal vitamin intake has been associated with a decreased risk for ASD particularly if taken before or near conception (Schmidt et al., 2011), as has folic acid intake (Schmidt et al., 2012; Surén et al., 2013; Wang, Li, Zhao, & Li, 2017). However, a recent study has suggested that very high maternal folate levels at birth are associated with a 2.5 times greater risk for ASD compared to folate levels at the 80$^{th}$ percentile, finding similar relationships with maternal Vitamin B12 levels and self-reported maternal multi-vitamin intake (Raghavan et al., 2018). Vitamin D and fish oil intake has been investigated for associations with ASD as well but no relationships have been identified (Lyall, Munger, O'Reilly, Santangelo, & Ascherio, 2013; Surén et al., 2013). Perhaps the most investigated maternal lifestyle factor is smoking, but with very mixed findings overall. Many studies have identified no association between maternal smoking and ASD (Kalkbrenner et al., 2012; H. J. Larsson et al., 2005; Lee et al., 2012; Maimburg & Vaeth, 2006; Tran et al., 2013), with some studies also identifying no association but trends in a protective direction (Bilder, Pinborough-Zimmerman, Miller, & McMahon, 2009; Burstyn, Sithole, & Zwaigenbaum, 2010). However, an equal amount of evidence points to increased risk for ASD with maternal

smoking (Hultman, Sparén, & Cnattingius, 2002; Hvidtjørn et al., 2011; M. Larsson, Weiss, Janson, Sundell, & Bornehag, 2009; Ronald, Happé, Dworzynski, Bolton, & Plomin, 2010; Visser et al., 2013). Finally, a much less studied risk factor has been alcohol exposure; 2 previous studies have reported no association with ASD (Eliasen et al., 2010; Singer et al., 2017) but lower maternal alcohol consumption in mothers of ASD children compared to controls has also been reported in two studies (Singer et al., 2017; Visser et al., 2013).

Maternal infections and autoimmune disorders have been among the more studied risk factors for ASD. Several studies based in European or US populations have demonstrated an association between viral or bacterial maternal infection during pregnancy and increased ASD risk (Atladóttir et al., 2010; Lee et al., 2015; Zerbo et al., 2015). A recent meta-analysis of 15 studies confirmed this association and noted an enhanced risk for infections requiring hospitalization (Jiang et al., 2016). Finally, a recent study in a habitually under-studied minority population of African ancestry demonstrated increased risk for ASD due to prenatal exposure to maternal fever (Brucato et al., 2017).

In addition, many environmental chemicals have also been investigated for their roles in ASD. Air pollution is perhaps one of the most robust factors positively associated with ASD. This broad category includes metals and solvents (Kalkbrenner et al., 2010; Roberts et al., 2013; Windham, Zhang, Gunier, Croen, & Grether, 2006), as well as particles, ozone, and nitrous oxides (Becerra, Wilhelm, Olsen, Cockburn, & Ritz, 2013; Kalkbrenner et al., 2010; Volk, Hertz-Picciotto, Delwiche, Lurmann, & McConnell, 2011; Volk, Lurmann, Penfold, Hertz-Picciotto, & McConnell, 2013; Windham et al., 2006).

#### *1.1.4 Genetic risk factors associated with ASD*

To date, most genetic factors associated with ASD risk have been rare *de novo* variants. For example, one landmark study utilized whole exome sequencing in more than 2,500 simplex families to find that *de novo* mutations (including both single nucleotide and copy number variants) account for 30% of all simplex cases and 45% of female cases of ASD, including likely gene disrupting (LGD) and missense events. Examination of the impact of LGD mutations on IQ level revealed a 5 point drop in IQ among males and a 20 point drop in IQ for recurrent LGD *de novo* mutations. Autism associated *de novo* mutations were also shown to be enriched in the gene classes of FMRP target genes, chromatin modifiers, and embryonically expressed genes (Iossifov et al., 2014). Studies have also demonstrated an increased burden of copy number variants (CNVs), including rare *de novo* CNVs, among individuals with autism (Sanders et al., 2015). In addition, a few studies have localized rare autism-associated CNVs and found the loci are also associated with developmental delay and schizophrenia (Sanders et al., 2015). Despite the differing locations and specific rare variants associated with ASD, the findings have been shown to converge on three biological pathways: Wnt signaling, chromatin remodeling, and synaptic function (Krumm, O'Roak, Shendure, & Eichler, 2014; Pinto et al., 2014).

As samples sizes for autism genome-wide association studies (GWAS) are starting to approach the numbers needed to achieve genome-wide significance, common risk variants for autism are beginning to emerge. The Psychiatric Genomics Consortium AUT workgroup has spearheaded the largest GWAS of autism to date. The first wave of results from the PGC-AUT group included 7,387 cases and 8,567 controls; they identified one genome-wide significant locus near *PITX3*, which is known to be involved in neuronal differentiation, and *CUEDC2*, previously associated with social skills (Autism Spectrum Disorders Working Group of The Psychiatric

Genomics Consortium, 2017). These results also confirmed previous finding showing a strong degree of genetic correlation between ASD and schizophrenia (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). The next wave of the PGC-AUT analysis included 18,381 ASD cases and 27,969 controls; five genome-wide significant loci were identified and results have recently been published as a preprint (Grove et al., 2017). As sample sizes continue to increase, more common variants as expected to be identified, as has been demonstrated in other psychiatric disorder GWAS (Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

Given that common variant findings for ASD are only now emerging, the functional biology implicated by these variants is not well understood. Moreover, many GWAS findings in general are often in inter-genic and/or non-coding areas of the genome, and thus lack an obvious functional consequence (Schierding, Cutfield, & O'Sullivan, 2014). This dissertation aims to elucidate the potential functional impact of implicated ASD common variation by integrating ASD GWAS results with epigenetic, specifically DNA methylation, data (**Chapter 2** and **Chapter 3**).

## 1.2 Epigenetics and ASD

### *1.2.1 Description of epigenetics*

Epigenetics is the study of reversible, mitotically heritable molecular information that does not include the DNA sequence itself. Epigenetic marks regulate transcription, imprinting, organismal development and tissue differentiation, and nuclear organization; they include histone modifications, small RNAs, and DNA methylation. This dissertation project focuses on the most

easily measured and thus most well studied epigenetic mark: DNA methylation (DNAm). This term is used to describe the covalent addition of a methyl or hydroxymethyl group to cytosine bases, and is a phenomenon that primarily occurs at CG dinucleotides in the mammalian genome, or CpG sites. CpG islands are characterized by dense clusters of CpG sites. They tend to by unmethylated and to be located in gene promoters. In general, most but not all, increases in CpG island methylation are associated with gene silencing (Illingworth & Bird, 2009). Unlike promoter and CpG island regions of the genome, DNA methylation located within the body of genes tends to be directly related to the expression of that gene, e.g. increased methylation is related to increased gene expression (Lou et al., 2014). DNAm at locations proximal (roughly a 2 Kb window) to CpG islands, known as CpG island shores, is often characterized by sites that are differentially methylated by tissue type and adverse health outcomes (Irizarry et al., 2009).

### *1.2.2 The relevance of epigenetics in ASD etiology*

There is compelling evidence suggesting that the etiology of ASD involves epigenetic mechanisms. First, certain disorders that overlap with ASD phenotypically such as Rett, Angelman, and Fragile X syndromes all have an epigenetic basis (Budimirovic & Kaufmann, 2011; Jedele, 2007; Mount, Charman, Hastings, Reilly, & Cass, 2003). Second, one of the major pathways on which rare variant findings for ASD converge is that of chromatin regulation or remodeling (Krumm et al., 2014; Pinto et al., 2014; Sanders et al., 2015). Most previous studies of ASD and epigenetic factors have examined DNAm, the most well understood and easily measured epigenetic mark.

A large proportion of prior human epigenetic studies of autism have examined DNAm measurements in postmortem brain tissue. Genome-wide approaches in brain tissue have

implicated various genes, such as *TSAPN32*/*C11orf21*, *ZFP57*, *PRRT1*, and *SDHAP3* (C. Ladd-Acosta et al., 2014; Nardone et al., 2014), while candidate gene approaches have found associations with the *MeCP2*, *OXTR*, *EN-2*, and *SHANK3* genes (Gregory et al., 2009; James, Shpyleva, Melnyk, Pavliv, & Pogribny, 2013; Nagarajan, Hogart, Gwye, Martin, & LaSalle, 2006; Zhu et al., 2014). Another recent brain-based study interrogated, on a genome-wide scale, another type of epigenetic mark, histone acetylation. Specifically analyzing the acetylation mark H3K27ac, the authors discovered large scale dysregulation of histone acetylation in the ASD cerebral cortex among a subset of all ASD cases they examined (Sun et al., 2016). Many studies on peripheral blood-based DNAm have also been conducted on a genome-wide scale – often using measurement platforms with less genomic coverage than used in the brain – and on a candidate gene scale as well. These analyses have also implicated various genes such as *NFYC*, *PTPRCAP*, and *MBD4* (Aldinger, Plummer, & Levitt, 2013; Gregory et al., 2009; Nguyen, Rauch, Pfeifer, & Hu, 2010; Wong et al., 2014). A recent preprint meta-analyzed blood-based DNAm from neonatal and early life blood samples, allowing for the largest EWAS for ASD to be performed to date. No genome-wide significant loci were found, though 45 loci passed a suggestive threshold for significance of 5E-5. Interestingly, the authors did discover significant associations between increased polygenic burden for ASD and two CpG sites near an ASD GWAS signal (Hannon et al., 2017).

Finally, there have been other peripheral tissues analyzed for DNAm relationships to ASD. One study used DNAm from buccal epithelium cells in ASD cases and controls all born to mothers at least 35 years of age. 15 DMRs in 14 genes were identified, many of which had been previously reported in the ASD CNV literature. These genes were also enriched in neuronal development

pathways (Berko et al., 2014). Another study using DNAm measured in sperm of fathers with a previously diagnosed ASD child identified 193 differentially methylated regions associated with the Autism Observational Score for Infants metric obtained at 12 months of age. These regions tended to cluster in genes involved in developmental processes (Feinberg et al., 2015).

### *1.2.3 Framework for understanding epigenetic involvement in ASD*

There are several pathways in which epigenetic mechanisms can be associated with ASD (**Figure 1.1**). Mechanistically, epigenetics can mediate genetic and/or environmental risk factors associated with ASD. In addition, epigenetics can illuminate gene-environment interactions. Environmental insults result in epigenetic changes, which, in turn, can provide different contexts in which genotype effects on phenotype may differ. Conversely, genotype can create different epigenetic states, which can dictate or influence how an environmental insult leads to disease. These constructs can inform disease etiology and intervention targets. It is also possible that epigenetic changes are a consequence of ASD as opposed to cause. If so, understanding these epigenetic changes could still be useful because they can serve as biomarkers of disease, which can help with disease diagnosis or treatment, and identification of at risk populations (Christine Ladd-Acosta & Fallin, 2016).

### *1.2.4 Challenges with studying epigenetics in autism: tissue availability*

While the most etiologically relevant tissue for mechanistic studies in ASD is brain tissue, we cannot measure brain methylation in living individuals. Therefore, analysis of this tissue type is limited to post-mortem samples which is not ideal because DNAm signatures to be analyzed may have only manifested after ASD onset (limiting etiologic inferences). Although organized efforts are underway to collect this type of tissue, there are an extremely limited number of postmortem

autism brain samples available at this time, further limiting our ability to examine the ASD brain epigenome. More accessible tissues that can be examined in a large number of living individuals include cord blood, placenta, saliva and buccal swabs, and peripheral blood tissues. However, it is unclear whether these accessible tissue sources can provide etiologically relevant results (Bakulski, Halladay, Hu, Mill, & Fallin, 2016). The degree to which blood can serve as a surrogate tissue for brain or other tissues is still being explored; while high levels of cross-tissue DNAm correlation at specific sites is not frequently observed, when concordance is observed, it is likely due to genetic influences (Davies et al., 2012; Hannon, Lunnon, Schalkwyk, & Mill, 2015).

This dissertation aims to understand the extent which these peripheral tissues can be used for DNAm studies of ASD, specifically focusing on DNAm sites that are associated with ASD-related SNPs. We integrate ASD GWAS findings with DNAm data from fetal brain, and compare SNP- and CpG-based inferences to the same integration from peripheral tissue such as cord blood and early-life peripheral blood (**Chapter 2**) and placenta (**Chapter 3**).

#### *1.2.5 Challenges in studying epigenetics in autism: measurement platforms*

The advent of commercially available, affordable, high-throughput, genome-wide array measurements for DNA methylation, such as the Illumina HumanMethylation450 BeadChip (450k Array), have led to the emergence of epigenome-wide association studies (EWAS). The 450k Array allows for the interrogation of roughly 485,000 CpG sites in the human genome, covering 99% of RefSeq genes and 96% of CpG islands (Bibikova et al., 2011). And, more recently, the next iteration of the 450k Array was released, known as the EPIC Array. This platform retains the balance of epigenome coverage and cost per sample as its predecessor, but

expands coverage to roughly 850,000 CpG sites, including more than 90% of those sites covered on the 450k Array. A main focus of this newer technology is the improved coverage of enhancer regions (Moran, Arribas, & Esteller, 2016; Pidsley et al., 2016). While array-based technologies continue to be necessary for methylome interrogation of epidemiologic-scale cohorts, they do suffer from a lack of comprehensive coverage of loci. Roughly 28 million CpG sites exist in the human genome; even the most comprehensive Illumina array platform only covers about 4% of methylation loci (Clark et al., 2012). Although the arrays are enriched for high value content and disease-relevant regions of the genome, it is still possible that disease-associated loci are not present on arrays and, thus, could be missed. An alternative methylation profiling technique known as whole genome bisulfite sequencing (WGBS) offers comprehensive, unbiased coverage of the epigenome. WGBS undoubtedly offers the greatest coverage compared to the widely used array technologies but is cost-prohibitive on large numbers of samples (Yong, Hsu, & Chen, 2016). This dissertation makes use of a unique collection of WGBS placenta samples to be able to perform an unbiased, comprehensive query of the DNA methylome for signatures related to ASD diagnosis. This study offers an examination of genome wide, sequencing-based DNAm and ASD, which has rarely been performed previously. However, sample sizes for individual, sex-stratified comparisons were limited. This dissertation also makes use of 450K-based DNAm from peripheral blood and cord blood samples to generate maps of methylation quantitative trait loci. While these data are inherently limited to the genomic regions queried by the 450K, they are produced via large numbers of samples and thus offer high confidence findings.

### *1.2.6 Importance of integrating genetic and epigenetic data in ASD.*

Genetic control of methylation has been demonstrated both in local (*cis*, SNP and CpG site on same chromosome) and more distant (*trans*, SNP and CpG site on different chromosomes) contexts. MeQTLs share properties of expression quantitative trait loci (eQTLs), though they are not perfectly overlapping. In fact, while meQTLs are enriched for eQTLs, only about 10% of eQTL relationships are completely mediated by DNAm. This reflects the fact that while DNAm plays a role in gene expression, other factors such as transcription factors (TFs), chromatin state, and nucleosome positioning are important contributors as well (Bell et al., 2011). MeQTLs have been shown to be enriched in various regulatory elements of the genome such as TFBS and DNase hypersensitive sites, across various tissue types (Hannon et al., 2016; McClay et al., 2015; Shi et al., 2014). Stronger SNP to CpG associations have been observed at smaller distances between SNP and CpG sites relative to associations at increasing distance (Hannon et al., 2016; Shi et al., 2014).

Recently, several studies have illustrated the potential for meQTLs to inform etiology of bipolar disorder (Gamazon et al., 2013) and schizophrenia (Hannon et al., 2016; van Eijk et al., 2015). These studies described an integration of meQTLs with GWAS summary statistics in what can be described as a "genotype-first" approach and a "methylation-first approach". In the "genotype-first" approach, disorder-associated SNPs are tested for their enrichment for meQTLs. This question aims to prove a global role for disorder-associated common variation with DNAm. The "methylation-first" approach undertaken by these studies is more centered on individual loci. The location of CpG site targets of disorder-associated SNPs can be used to expand SNP-based loci beyond the locations identified by SNPs alone. These expanded loci can then indicate additional genes warranting functional follow up, as they can be putative causal loci for the

phenotype (Hannon et al., 2016; van Eijk et al., 2015). Given the success of these techniques in bipolar disorder and schizophrenia, which have well-established genetic overlap with autism spectrum disorder (ASD), and the emerging evidence for an epigenetic basis to ASD, integration of genetic and epigenetic data in an ASD context is warranted. We exploit these same methodologies, adding an additional "methylation-first" approach that seeks to understand the biological pathways enumerated by ASD SNP-controlled CpG sites. Our studies are undertaken across fetal brain, cord blood, peripheral childhood blood, and placenta tissues to provide cross-tissue insights into ASD SNP-methylation relationships and etiology (**Chapter 2** and **Chapter 3**).

## 1.3 The placenta

### *1.3.1 General structure and function*

The placenta is a unique structure formed during pregnancy that consists of a maternal contribution (known as the basal plate) and a fetal contribution (chorionic plate). The placenta acts as a maternal-fetal interface that allows for nutrient uptake, waste elimination, and gas exchange. The basal plate, derived from the endometrium, consists of a layer of epithelial cells and blood vessels that extend from the uterine wall. The chorionic plate, derived from the chorionic sac, is anchored by the umbilical cord that extends from the fetus. The umbilical cord contains two arteries, which take deoxygenated blood away from the fetus, and one vein which brings oxygenated blood towards the fetus. These blood vessels protrude towards the maternal side in what is known as the intervillous space, and branch out in structures known as villi (to allow for increased surface area). Here, maternal blood that enters the intervillous space via

spiral endometrial arteries and constantly bathe the villous structures to promote maternal-fetal nutrient exchange and waste elimination (Gude, Roberts, Kalionis, & King, 2004).

The growth and maintenance of placenta function during pregnancy involves a highly dynamic but tightly regulated process of cellular differentiation. Placental cytotrophoblasts can be broadly dichotomized into those that differentiate towards the villous and the extravillous pathways. The placental villi contain an inner cytotrophoblasts layer and an outer syncytiotrophoblast layer, which is a continuous, multinucleated mass of cells that is the only actual fetal portion of the placenta that makes contact with maternal blood. Another group of cytotrophoblasts operate outside of the villous structures and serve other functions, such as the remodeling of uterine arteries and tunneling of the uterine wall.

### *1.3.2. The role of the placenta in fetal brain development*

The current framework linking placenta function to increased susceptibility to neurodevelopmental disorders involves the placenta's role in regulating maternal-fetal interactions. Maternal insults such as immune activation, malnutrition, or stress, are thought to lead to altered placenta status in the state of cytokines, hormones and endocrine factors, and gene expression. These altered states then disrupt placental processes such as immune tolerance, the growth of the placenta, invasion into the uterine wall, and the production of neuroactive molecules. Finally, these affected processes result in several adverse events linked to the prenatal programming of disease, such as immune dysfunction, intrauterine growth restriction or malnutrition, hypoxia, ischemia, and altered neuronal development and circuit formation, each of which contribute to an increased susceptibility to neurodevelopmental disorders (Hsiao & Patterson, 2012).

There has been a wave of recent work relating placenta epigenetic features to neurodevelopmental outcomes, broadly. There has been one previous study specifically examining the ASD phenotype for distinguishing placenta DNAm profiles. This study was successful in identifying one autosomal differentially methylated region in *DLL1* (Schroeder et al., 2016). Numerous other studies, most via a candidate gene approach but two via a genome-wide approach, have investigated infant-based neurobehavioral outcomes with placenta DNAm (Lester & Marsit, 2018). For example, one of the genome-wide studies found 2 genome-wide significant loci associated with infant attention harbored in *FHIT* and *ANKRD11*. The authors examined 4 domains of infant behavior in addition to attention: arousal, lethargy, and quality of movement, finding that suggestively associated CpG sites were almost always unique to each outcome. However, at a gene level, 51 genes were differentially methylated (again at suggestive levels of significance) to some extent in all 4 outcomes; gene ontology analyses indicated that these genes were involved in pathways such as cellular adhesion and nervous system development (Paquette et al., 2016).

In addition, altered placenta structure and/or function has also been directly linked to ASD. One study based in the prospective, enriched risk birth cohort Markers of Autism Risk in Babies – Learning Early Signs (MARBLES), demonstrated that mothers with a previous child with diagnosed ASD had an 8-fold greater odds of having placentas (from a subsequent pregnancy) with two or more trophoblast inclusions compared to control mothers (Walker et al., 2013). Trophoblast inclusions are a hallmark of abnormal growth, and are characterized by the typically inner cytotrophoblast cell layer enveloping the typically outer syncytiotrophoblast layer (Adler, Madankumar, Rosner, & Reznik, 2014). It is hypothesized that increased serotonin concentration

*in utero*, which itself is independently linked to ASD (Mulder et al., 2004), leads to increased cytotrophoblast proliferation, which then in turn can lead to trophoblast inclusions (Fecteau & Eiler, 2001; Walker et al., 2013). The MARBLES-based study also identified mothers with ASD as having a 5.5-fold higher odds of having a calcified trophoblast inclusion compared to control mothers; calcification in trophoblast inclusions is thought to be a mark of age and impede blood flow (Walker et al., 2013). In addition to this study examining pregnancies subsequent to the ASD-affected child, similar relationships demonstrating an association between trophoblast inclusions and ASD have also come in directly comparing placentas from children who went on to develop ASD to those from control children (Anderson, Jacobs-Stannard, Chawarska, Volkmar, & Kliman, 2007). Another more recent study also linked various aspects of placenta histopathology to ASD. Specifically, acute placenta inflammation, chronic uteroplacental vasculitis, the fetal inflammatory response in the chorionic plate, and maternal vascular malperfusion pathology were all associated with placentas from children who went on to develop ASD compared to those from matched controls (Straughen et al., 2017). Finally, another paper demonstrated the utility of random forest-based classification of placenta chorionic surface vascular network structure to distinguish between placentas from high risk ASD pregnancies (i.e. those from mothers with a child previously diagnosed with ASD) and population-based controls (Chang et al., 2017).

#### *1.3.3. Differences in disease prevalence by sex and sexual dimorphism in the placenta*

It is well established that there are sex biases in phenotypes across the life course. For example, males are 20% more likely to have a poorer outcome in pregnancies complicated by preeclampsia, preterm delivery, and intrauterine growth restriction (IUGR) (Vatten & Skjaerven,

2004). And later in life, males are more likely to be diagnosed with autism spectrum disorder (ASD) (Christensen et al., 2016), schizophrenia (Aleman, Kahn, & Selten, 2003), and cardiovascular outcomes (Bots, Peters, & Woodward, 2017), while females are more likely to be diagnosed with autoimmune disorders such as systemic lupus erythematosus, Sjogren's syndrome, and rheumatoid arthritis (Ngo, Steyn, & McCombe, 2014). Moreover, it is increasingly appreciated that the developmental and/or gestational period is critical to the programming of disease late in life – the "developmental origins of health and disease" construct (Wadhwa, Buss, Entringer, & Swanson, 2009). The placenta is the key organ at play during fetal development, providing a critical interface between the mother and fetus and through regulating the transfer of nutrients, waste, and oxygen (Gude et al., 2004). Taken together, these notions suggest that there are sex-specific placental functions that can ultimately contribute to skewed sex ratios for adverse health outcomes early and later in life.

Male-female differences have been demonstrated during the gestational period, particularly in their adaptations to adverse insults or environments (Clifton, 2010). For example, female growth has been shown to be reduced in the presence of mild maternal asthma; growth can be returned to normal levels with use of inhaled steroids, however (Murphy et al., 2003). Alternatively males continue normal growth in the presence of maternal asthma and only exhibit signs of affected growth when the asthma reaches acute levels (Murphy, Gibson, Talbot, & Clifton, 2005). Similarly, female fetuses in pregnancies complicated by preeclampsia have shown to exhibit reduced growth, while males tend to grow normally (Stark, Clifton, & Wright, 2009). Such findings have led to hypotheses regarding the differential nature of responses to perturbations during pregnancy by sex: that males aim to maintain growth in the presence of an initial adverse

insult, leaving less reserve capacity to handle an additional adverse event. In contrast, females reduce growth in the presence of a poor intrauterine environment, allowing for increased ability to address further adverse events (Clifton, 2010). It is also postulated that females tend to be more at risk for adverse outcomes when perturbations occur in the peri-conception or early gestational periods, affecting trophoblast differentiation. In contrast, male fetuses are more vulnerable in mid-to late gestation, when exposures will affect expansion and function of the definitive placenta and affect nutrient transport (Kalisch-Smith, Simmons, Dickinson, & Moritz, 2017). Finally, placenta immune function displays evidence for sex-specific effects as well; for example mouse studies have shown maternal obesity to be linked to placenta inflammation in late gestation, with males displaying greater inflammation and macrophage activation than females (Kim, Young, Grattan, & Jasoni, 2014; Tarrade, Panchenko, Junien, & Gabory, 2015).

However, little is understood with respect to the mechanism of sex differences in the placenta. Some gene expression studies have been conducted; for example a study investigating sorted placenta cell types found pathways such as graft vs. host disease, immune function, and inflammation to be enriched in genes differentially expressed by sex (Cvitic et al., 2013). Another meta-analysis of placenta microarray data discovered pathways such as cell growth, proliferation, and hormonal function (Buckberry, Bianco-Miotto, Bent, Dekker, & Roberts, 2014). Sex-related mechanisms may be additionally illuminated through the placenta methylome. Specifically, understanding the mechanisms of placenta structural or functional differences by sex can thus shed light on the mechanisms leading to highly sex-biased phenotypes such as ASD.

### *1.3.4. Placenta Epigenetics*

The placental epigenome is of increasing area of interest both in the context of normal development and disease states. A seminal paper characterizing the placenta methylome demonstrated its unique tendency to contain partially methylated domains (PMDs), which are associated with gene repression and inactive chromatin. The authors discovered some additional features such as the fact that genes within placenta PMDs have tissue-specific functions, and that in regulatory regions, methylation levels in promoter CpG islands are higher for genes within PMDs. Additionally, PMDs are broadly of interest outside of the context of the placenta specifically. For example, the authors found that genes harbored in regions that are PMDs in placenta, SH-SY5Y (neuronal), and IMR90 (lung fibroblast), are enriched for pathways such as sensory perception of smell, neurological system processes, cell surface receptor signals, and keratinization. Finally, PMDs contain many neuronal development genes, making their investigation in an ASD context a promising endeavor (Schroeder et al., 2013). This study was carried out by the same authors some years later (see 'Epigenetics of ASD'), with good success (Schroeder et al., 2016).

There have been many other successful studies identifying placental DNAm changes associated with many other phenotypes such as preeclampsia (Kulkarni, Chavan-Gautam, Mehendale, Yadav, & Joshi, 2011; Yuen, Peñaherrera, von Dadelszen, McFadden, & Robinson, 2010) and IUGR (Roifman et al., 2016). Another branch of studies has established the sensitivity of the placenta methylome to various environmental exposures. For example two previous studies, both using an early iteration of the Illumina methylation array technology (27K Array), have identified CpG sites significantly associated with prenatal smoking (Maccani, Koestler, Houseman, Marsit, & Kelsey, 2013; Suter et al., 2011). Two more recent studies, both using a

later iteration of Illumina methylation arrays (450K Array), also identified placenta methylation sites significantly associated with mercury (Maccani et al., 2015) and cadmium (Mohanty et al., 2015). The degree of genetic control of placenta methylation signals has not been investigated on a genome-wide scale. **Chapter 3** of this dissertation addresses this gap by defining, to our knowledge, the first comprehensive meQTL map in placenta tissue and using this map to provide insights into ASD biology. Another key feature of the placenta methylome is its dynamic nature across the gestational period. Specifically, it has been demonstrated that global methylation levels tend to rise with increasing gestational age (Novakovic et al., 2011). Therefore, association analyses must consider gestational age as a potential confounding factor, as gestational age can be linked to various phenotypes as well. Moreover, studies from term placenta samples must take caution to claim functional implications of any significant findings, as the identified DNAm signatures may not be reflective of the DNAm state during the gestational period.

Finally, two previous studies have investigated the placenta methylome in the context of sexual dimorphism. These efforts are buoyed by many successful efforts to identify autosomal sex-specific DNAm differences in tissues such as peripheral blood (Inoshita et al., 2015; Price et al., 2013; Singmann et al., 2015; Suderman et al., 2017), cord blood (Yousefi et al., 2015), and prefrontal cortex (Xu et al., 2014). One the previous placenta studies successfully identified 91 significant autosomal sites related to fetal sex in a small for gestational age cohort, implicating similar pathways to those found by gene expression studies such as immune function, growth/transcription factor signaling, and transport across cell membranes. However, this previous study utilized Illumina 450k microarrays for DNAm measurement (Martin et al., 2017).

While the standard measurement tool for DNAm studies, this platform is still limited in its coverage of the DNA methylome. The second placenta study addressed these concerns by measuring placenta DNAm via WGBS. This study identified many sex chromosome DNAm differences by sex, which is expected, but also one autosomal differentially methylated region in the gene body of *CSMD1*. The authors also failed to identify this sex DMR in 8 other tissues, suggesting that it is placenta-specific. One limitation of this study was that sample sizes were limited to 2 male and female samples each (Gong et al., 2018). In **Chapter 5** of this dissertation, we extend this area of research by conducting a whole-genome bisulfite study for sex-specific DNAm patterns in placenta tissue from a substantial number of samples ($N_{males}$ = 17, $N_{females}$ = 20).

### 1.4. Focus of this dissertation

This dissertation is centered on gaining etiologic insights towards ASD and sexual dimorphism during pregnancy through epigenetic data, particularly through its integration with genotype data, across tissue types. With respect to ASD, this work can aid our understanding of the genetic and epigenetic dysregulation characteristic of ASD, provide intervention targets, and provide mechanistic insights to observed genetic associations. Moreover, this work can shed light on specific placental functions that differ in normal development according to fetal sex, allowing for an understanding of how the gestational period can program sex-biased phenotypes, like ASD, throughout the life course.

Our first goal is to better characterize the functional biology of common variants that are increasingly being identified in ASD. Given the growing evidence for an epigenetic basis to ASD, we pursue this goal by incorporating epigenetic data, specifically DNA methylation, along

with ASD GWAS association data. To simultaneously investigate the utility of peripheral tissues for DNAm-based insights into a brain-based disease such as ASD, we conduct these genotype-epigenotype integrations using DNAm data from fetal brain, peripheral blood, and cord blood (**Chapter 2**), and additionally placenta (**Chapter 3**). Specifically, we determined whether ASD associated SNPs were enriched for meQTLs and also whether particular pathways were enriched in the methylation targets of ASD SNPs. We compare SNP and CpG-based insights from peripheral tissues to those from brain, to better understand the suitability of various tissues for etiologic insights into ASD. These insights can also reveal tissue-specific functionality of GWAS SNPs that better characterizes how their effect on ASD is context-dependent.

Our next goal is to search for ASD-related signatures of placenta DNAm. Most previous studies of ASD and DNAm have taken place using brain samples and peripheral tissues such as blood. While brain-based DNAm is of most interest for etiologic insights into a neurodevelopmental disorder such as ASD, accessibility is limited to post-mortem samples. Peripheral tissues are inherently more accessible from brain, but are less useful for identifying mechanistic etiologic insights into ASD. The placenta is a unique peripheral tissue in an ASD context because it is similarly accessible, but more connected to ASD etiology. Specifically, genetic factors and environmental insults during pregnancy can act through placenta DNAm to influence placenta structure and function, which in turn influences fetal brain development. Despite this compelling framework and rationale for placenta DNAm and ASD, only one previous study has examined this relationship. Therefore, we undertake a study searching for differentially methylated regions and global methylation differences in placenta related to ASD diagnosis (**Chapter 4**).

Finally, the role of placenta DNAm in contributing to phenotype is one in which genetic and environmental insults shape DNAm, which in turn effects placenta structure and function, which then leads to health outcomes. This framework fits squarely within the realm of the developmental origins of health and disease construct, which suggest that early life factors program disease later in life (Bianco-Miotto, Craig, Gasser, van Dijk, & Ozanne, 2017). Taken together with the fact that certain phenotypes, such as ASD, maintain a highly sex-biased prevalence, it follows logically that this placenta framework itself is sex-specific. Indeed, evidence exists for the differential response of the male and female fetus to certain insults during the gestational period, such as asthma (Murphy et al., 2003, 2005). But little is known with respect to the mechanisms of these sex-specific placental responses and functions. Therefore, we perform an epigenome-wide association study searching for differentially methylated regions associated with fetal sex. Additional follow-up analyses of a fetal sex-related region incorporate data from other tissue types to assess placental tissue specificity, and integrate genotype data to assess epigenetic mediation of genetic variation, at both the local and distal level, on placenta morphological features (**Chapter 5**).

The work presented in this dissertation demonstrates the utility of integrating genetic and epigenetic data across tissue types to extend research on the genetic and epigenetic etiology of ASD, and for uncovering the mechanisms of sex-specific placental function and the potential programming of sex-biased disease later in life.

### 1.5. References

Adler, E., Madankumar, R., Rosner, M., & Reznik, S. E. (2014). Increased placental trophoblast inclusions in placenta accreta. *Placenta*, *35*(12), 1075–1078. https://doi.org/10.1016/j.placenta.2014.09.014

Aldinger, K. A., Plummer, J. T., & Levitt, P. (2013). Comparative DNA methylation among females with neurodevelopmental disorders and seizures identifies TAC1 as a MeCP2 target gene. *Journal of Neurodevelopmental Disorders*, *5*(1), 15. https://doi.org/10.1186/1866-1955-5-15

Aleman, A., Kahn, R. S., & Selten, J.-P. (2003). Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Archives of General Psychiatry*, *60*(6), 565–571. https://doi.org/10.1001/archpsyc.60.6.565

Anderson, G. M., Jacobs-Stannard, A., Chawarska, K., Volkmar, F. R., & Kliman, H. J. (2007). Placental trophoblast inclusions in autism spectrum disorder. *Biological Psychiatry*, *61*(4), 487–491. https://doi.org/10.1016/j.biopsych.2006.03.068

Atladóttir, H. O., Thorsen, P., Østergaard, L., Schendel, D. E., Lemcke, S., Abdallah, M., & Parner, E. T. (2010). Maternal infection requiring hospitalization during pregnancy and autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(12), 1423–1430. https://doi.org/10.1007/s10803-010-1006-y

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism*, *8*, 21. https://doi.org/10.1186/s13229-017-0137-9

Bakulski, K. M., Halladay, A., Hu, V. W., Mill, J., & Fallin, M. D. (2016). Epigenetic Research in Neuropsychiatric Disorders: the "Tissue Issue." *Current Behavioral Neuroscience Reports*, *3*(3), 264–274. https://doi.org/10.1007/s40473-016-0083-4

Becerra, T. A., Wilhelm, M., Olsen, J., Cockburn, M., & Ritz, B. (2013). Ambient air pollution and autism in Los Angeles county, California. *Environmental Health Perspectives*, *121*(3), 380–386. https://doi.org/10.1289/ehp.1205827

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., … Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology*, *12*(1), R10. https://doi.org/10.1186/gb-2011-12-1-r10

Berko, E. R., Suzuki, M., Beren, F., Lemetre, C., Alaimo, C. M., Calder, R. B., … Greally, J. M. (2014). Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genetics*, *10*(5), e1004402. https://doi.org/10.1371/journal.pgen.1004402

Bianco-Miotto, T., Craig, J. M., Gasser, Y. P., van Dijk, S. J., & Ozanne, S. E. (2017). Epigenetics and DOHaD: from basics to birth and beyond. *Journal of Developmental Origins of Health and Disease*, *8*(5), 513–519. https://doi.org/10.1017/S2040174417000733

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., … Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bilder, D., Pinborough-Zimmerman, J., Miller, J., & McMahon, W. (2009). Prenatal, perinatal, and neonatal factors associated with autism spectrum disorders. *Pediatrics*, *123*(5), 1293–1300. https://doi.org/10.1542/peds.2008-0927

Bots, S. H., Peters, S. A. E., & Woodward, M. (2017). Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health*, *2*(2), e000298. https://doi.org/10.1136/bmjgh-2017-000298

Brucato, M., Ladd-Acosta, C., Li, M., Caruso, D., Hong, X., Kaczaniuk, J., … Wang, X. (2017). Prenatal exposure to fever is associated with autism spectrum disorder in the boston birth cohort. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(11), 1878–1890. https://doi.org/10.1002/aur.1841

Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A., & Roberts, C. T. (2014). Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. *Molecular Human Reproduction*, *20*(8), 810–819. https://doi.org/10.1093/molehr/gau035

Budimirovic, D. B., & Kaufmann, W. E. (2011). What can we learn about autism from studying fragile X syndrome? *Developmental Neuroscience*, *33*(5), 379–394. https://doi.org/10.1159/000330213

Burstyn, I., Sithole, F., & Zwaigenbaum, L. (2010). Autism spectrum disorders, maternal characteristics and obstetric complications among singletons born in Alberta, Canada. *Chronic Diseases in Canada*, *30*(4), 125–134.

Chang, J.-M., Zeng, H., Han, R., Chang, Y.-M., Shah, R., Salafia, C. M., … Croen, L. (2017). Autism risk classification using placental chorionic surface vascular network features. *BMC Medical Informatics and Decision Making*, *17*(1), 162. https://doi.org/10.1186/s12911-017-0564-8

Cheslack-Postava, K., Liu, K., & Bearman, P. S. (2011). Closely spaced pregnancies are associated with increased odds of autism in California sibling births. *Pediatrics*, *127*(2), 246–253. https://doi.org/10.1542/peds.2010-2371

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Centers for Disease Control and Prevention (CDC). (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years--Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly*

*Report. Surveillance Summaries (Washington, D.C.: 2002)*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Clark, C., Palta, P., Joyce, C. J., Scott, C., Grundberg, E., Deloukas, P., … Coffey, A. J. (2012). A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling. *PloS One*, *7*(11), e50233. https://doi.org/10.1371/journal.pone.0050233

Clifton, V. L. (2010). Review: Sex and the human placenta: mediating differential strategies of fetal growth and survival. *Placenta*, *31 Suppl*, S33-39. https://doi.org/10.1016/j.placenta.2009.11.010

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., … Bolton, P. (2015). Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry*, *72*(5), 415–423. https://doi.org/10.1001/jamapsychiatry.2014.3028

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., … International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994. https://doi.org/10.1038/ng.2711

Cvitic, S., Longtine, M. S., Hackl, H., Wagner, K., Nelson, M. D., Desoye, G., & Hiden, U. (2013). The human placental sexome differs between trophoblast epithelium and villous vessel endothelium. *PloS One*, *8*(10), e79233. https://doi.org/10.1371/journal.pone.0079233

Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., … Mill, J. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biology*, *13*(6), R43. https://doi.org/10.1186/gb-2012-13-6-r43

Eliasen, M., Tolstrup, J. S., Nybo Andersen, A.-M., Grønbaek, M., Olsen, J., & Strandberg-Larsen, K. (2010). Prenatal alcohol exposure and autistic spectrum disorders--a population-based prospective study of 80,552 children and their mothers. *International Journal of Epidemiology*, *39*(4), 1074–1081. https://doi.org/10.1093/ije/dyq056

Fecteau, K. A., & Eiler, H. (2001). Placenta detachment: unexpected high concentrations of 5-hydroxytryptamine (serotonin) in fetal blood and its mitogenic effect on placental cells in bovine. *Placenta*, *22*(1), 103–110. https://doi.org/10.1053/plac.2000.0596

Feinberg, J. I., Bakulski, K. M., Jaffe, A. E., Tryggvadottir, R., Brown, S. C., Goldman, L. R., … Feinberg, A. P. (2015). Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *International Journal of Epidemiology*, *44*(4), 1199–1210. https://doi.org/10.1093/ije/dyv028

Gamazon, E. R., Badner, J. A., Cheng, L., Zhang, C., Zhang, D., Cox, N. J., … Liu, C. (2013). Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Molecular Psychiatry*, *18*(3), 340–346. https://doi.org/10.1038/mp.2011.174

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Gillberg, C., & Fernell, E. (2014). Autism plus versus autism pure. *Journal of Autism and Developmental Disorders*, *44*(12), 3274–3276. https://doi.org/10.1007/s10803-014-2163-1

Gong, S., Johnson, M. D., Dopierala, J., Gaccioli, F., Sovio, U., Constância, M., … Charnock-Jones, D. S. (2018). Genome-wide oxidative bisulfite sequencing identifies sex-specific methylation differences in the human placenta. *Epigenetics*, 1–32. https://doi.org/10.1080/15592294.2018.1429857

Gregory, S. G., Connelly, J. J., Towers, A. J., Johnson, J., Biscocho, D., Markunas, C. A., … Pericak-Vance, M. A. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Medicine*, *7*, 62. https://doi.org/10.1186/1741-7015-7-62

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R., Won, H., … Børglum, A. D. (2017). Common risk variants identified in autism spectrum disorder. *bioRxiv*, 224774. https://doi.org/10.1101/224774

Gude, N. M., Roberts, C. T., Kalionis, B., & King, R. G. (2004). Growth and function of the normal human placenta. *Thrombosis Research*, *114*(5–6), 397–407. https://doi.org/10.1016/j.thromres.2004.06.038

Halladay, A. K., Bishop, S., Constantino, J. N., Daniels, A. M., Koenig, K., Palmer, K., … Szatmari, P. (2015). Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Molecular Autism*, *6*, 36. https://doi.org/10.1186/s13229-015-0019-y

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., … Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*, *68*(11), 1095–1102. https://doi.org/10.1001/archgenpsychiatry.2011.76

Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*, *10*(11), 1024–1032. https://doi.org/10.1080/15592294.2015.1100786

Hannon, E., Schendel, D., Ladd-Acosta, C., Grove, J., Group, iPSYCH-B. A., Hansen, C. S., … Mill, J. (2017). Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *bioRxiv*, 225193. https://doi.org/10.1101/225193

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., … Mill, J. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nature Neuroscience*, *19*(1), 48–54. https://doi.org/10.1038/nn.4182

Hansen, S. N., Schendel, D. E., & Parner, E. T. (2015). Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. *JAMA Pediatrics*, *169*(1), 56–62. https://doi.org/10.1001/jamapediatrics.2014.1893

Hsiao, E. Y., & Patterson, P. H. (2012). Placental regulation of maternal-fetal interactions and brain development. *Developmental Neurobiology*, *72*(10), 1317–1326. https://doi.org/10.1002/dneu.22045

Hultman, C. M., Sparén, P., & Cnattingius, S. (2002). Perinatal risk factors for infantile autism. *Epidemiology (Cambridge, Mass.)*, *13*(4), 417–423.

Hvidtjørn, D., Grove, J., Schendel, D., Schieve, L. A., Sværke, C., Ernst, E., & Thorsen, P. (2011). Risk of autism spectrum disorders in children born after assisted conception: a population-based follow-up study. *Journal of Epidemiology and Community Health*, *65*(6), 497–502. https://doi.org/10.1136/jech.2009.093823

Illingworth, R. S., & Bird, A. P. (2009). CpG islands--'a rough guide'. *FEBS Letters*, *583*(11), 1713–1720. https://doi.org/10.1016/j.febslet.2009.04.012

Inoshita, M., Numata, S., Tajima, A., Kinoshita, M., Umehara, H., Yamamori, H., … Ohmori, T. (2015). Sex differences of leukocytes DNA methylation adjusted for estimated cellular proportions. *Biology of Sex Differences*, *6*, 11. https://doi.org/10.1186/s13293-015-0029-7

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216–221. https://doi.org/10.1038/nature13908

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., … Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, *41*(2), 178–186. https://doi.org/10.1038/ng.298

Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., … Eichler, E. E. (2014). A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *American Journal of Human Genetics*, *94*(3), 415–425. https://doi.org/10.1016/j.ajhg.2014.02.001

James, S. J., Shpyleva, S., Melnyk, S., Pavliv, O., & Pogribny, I. P. (2013). Complex epigenetic regulation of engrailed-2 (EN-2) homeobox gene in the autism cerebellum. *Translational Psychiatry*, *3*, e232. https://doi.org/10.1038/tp.2013.8

Jedele, K. B. (2007). The overlapping spectrum of rett and angelman syndromes: a clinical review. *Seminars in Pediatric Neurology*, *14*(3), 108–117. https://doi.org/10.1016/j.spen.2007.07.002

Jiang, H.-Y., Xu, L.-L., Shao, L., Xia, R.-M., Yu, Z.-H., Ling, Z.-X., … Ruan, B. (2016). Maternal infection during pregnancy and risk of autism spectrum disorders: A systematic review and meta-analysis. *Brain, Behavior, and Immunity*, *58*, 165–172. https://doi.org/10.1016/j.bbi.2016.06.005

Kalisch-Smith, J. I., Simmons, D. G., Dickinson, H., & Moritz, K. M. (2017). Review: Sexual dimorphism in the formation, function and adaptation of the placenta. *Placenta*, *54*, 10–16. https://doi.org/10.1016/j.placenta.2016.12.008

Kalkbrenner, A. E., Braun, J. M., Durkin, M. S., Maenner, M. J., Cunniff, C., Lee, L.-C., … Daniels, J. L. (2012). Maternal smoking during pregnancy and the prevalence of autism spectrum disorders, using data from the autism and developmental disabilities monitoring network. *Environmental Health Perspectives*, *120*(7), 1042–1048. https://doi.org/10.1289/ehp.1104556

Kalkbrenner, A. E., Daniels, J. L., Chen, J.-C., Poole, C., Emch, M., & Morrissey, J. (2010). Perinatal exposure to hazardous air pollutants and autism spectrum disorders at age 8. *Epidemiology (Cambridge, Mass.)*, *21*(5), 631–641. https://doi.org/10.1097/EDE.0b013e3181e65d76

Kim, D. W., Young, S. L., Grattan, D. R., & Jasoni, C. L. (2014). Obesity during pregnancy disrupts placental morphology, cell proliferation, and inflammation in a sex-specific manner across gestation in the mouse. *Biology of Reproduction*, *90*(6), 130. https://doi.org/10.1095/biolreprod.113.117259

Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences*, *37*(2), 95–105. https://doi.org/10.1016/j.tins.2013.11.005

Kulkarni, A., Chavan-Gautam, P., Mehendale, S., Yadav, H., & Joshi, S. (2011). Global DNA methylation patterns in placenta and its association with maternal hypertension in pre-eclampsia. *DNA and Cell Biology*, *30*(2), 79–84. https://doi.org/10.1089/dna.2010.1084

Ladd-Acosta, C., & Fallin, M. D. (2016). The role of epigenetics in genetic and environmental epidemiology. *Epigenomics*, *8*(2), 271–283. https://doi.org/10.2217/epi.15.102

Ladd-Acosta, C., Hansen, K. D., Briem, E., Fallin, M. D., Kaufmann, W. E., & Feinberg, A. P. (2014). Common DNA methylation alterations in multiple brain regions in autism. *Molecular Psychiatry*, *19*(8), 862–871. https://doi.org/10.1038/mp.2013.114

Larsson, H. J., Eaton, W. W., Madsen, K. M., Vestergaard, M., Olesen, A. V., Agerbo, E., … Mortensen, P. B. (2005). Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. *American Journal of Epidemiology*, *161*(10), 916-925-928. https://doi.org/10.1093/aje/kwi123

Larsson, M., Weiss, B., Janson, S., Sundell, J., & Bornehag, C.-G. (2009). Associations between indoor environmental factors and parental-reported autistic spectrum disorders in children 6-8 years of age. *Neurotoxicology*, *30*(5), 822–831. https://doi.org/10.1016/j.neuro.2009.01.011

Lee, B. K., Gardner, R. M., Dal, H., Svensson, A., Galanti, M. R., Rai, D., … Magnusson, C. (2012). Brief report: maternal smoking during pregnancy and autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*(9), 2000–2005. https://doi.org/10.1007/s10803-011-1425-4

Lee, B. K., Magnusson, C., Gardner, R. M., Blomström, Å., Newschaffer, C. J., Burstyn, I., … Dalman, C. (2015). Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. *Brain, Behavior, and Immunity*, *44*, 100–105. https://doi.org/10.1016/j.bbi.2014.09.001

Lester, B. M., & Marsit, C. J. (2018). Epigenetic mechanisms in the placenta related to infant neurodevelopment. *Epigenomics*. https://doi.org/10.2217/epi-2016-0171

Lou, S., Lee, H.-M., Qin, H., Li, J.-W., Gao, Z., Liu, X., … Yip, K. Y. (2014). Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. *Genome Biology*, *15*(7), 408. https://doi.org/10.1186/s13059-014-0408-0

Lyall, K., Munger, K. L., O'Reilly, É. J., Santangelo, S. L., & Ascherio, A. (2013). Maternal dietary fat intake in association with autism spectrum disorders. *American Journal of Epidemiology*, *178*(2), 209–220. https://doi.org/10.1093/aje/kws433

Lyall, K., Schmidt, R. J., & Hertz-Picciotto, I. (2014). Maternal lifestyle and environmental risk factors for autism spectrum disorders. *International Journal of Epidemiology*, *43*(2), 443–464. https://doi.org/10.1093/ije/dyt282

Maccani, J. Z. J., Koestler, D. C., Houseman, E. A., Marsit, C. J., & Kelsey, K. T. (2013). Placental DNA methylation alterations associated with maternal tobacco smoking at the RUNX3 gene are also associated with gestational age. *Epigenomics*, *5*(6), 619–630. https://doi.org/10.2217/epi.13.63

Maccani, J. Z. J., Koestler, D. C., Lester, B., Houseman, E. A., Armstrong, D. A., Kelsey, K. T., & Marsit, C. J. (2015). Placental DNA Methylation Related to Both Infant Toenail Mercury and Adverse Neurobehavioral Outcomes. *Environmental Health Perspectives*, *123*(7), 723–729. https://doi.org/10.1289/ehp.1408561

Maimburg, R. D., & Vaeth, M. (2006). Perinatal risk factors and infantile autism. *Acta Psychiatrica Scandinavica*, *114*(4), 257–264. https://doi.org/10.1111/j.1600-0447.2006.00805.x

Martin, E., Smeester, L., Bommarito, P. A., Grace, M. R., Boggess, K., Kuban, K., … Fry, R. C. (2017). Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period. *Epigenomics*, *9*(3), 267–278. https://doi.org/10.2217/epi-2016-0132

McClay, J. L., Shabalin, A. A., Dozmorov, M. G., Adkins, D. E., Kumar, G., Nerella, S., … van den Oord, E. J. C. G. (2015). High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biology*, *16*, 291. https://doi.org/10.1186/s13059-015-0842-7

Mohanty, A. F., Farin, F. M., Bammler, T. K., MacDonald, J. W., Afsharinejad, Z., Burbacher, T. M., … Enquobahrie, D. A. (2015). Infant sex-specific placental cadmium and DNA methylation associations. *Environmental Research*, *138*, 74–81. https://doi.org/10.1016/j.envres.2015.02.004

Moran, S., Arribas, C., & Esteller, M. (2016). Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*, *8*(3), 389–399. https://doi.org/10.2217/epi.15.114

Mount, R. H., Charman, T., Hastings, R. P., Reilly, S., & Cass, H. (2003). Features of autism in Rett syndrome and severe mental retardation. *Journal of Autism and Developmental Disorders*, *33*(4), 435–442.

Mulder, E. J., Anderson, G. M., Kema, I. P., de Bildt, A., van Lang, N. D. J., den Boer, J. A., & Minderaa, R. B. (2004). Platelet serotonin levels in pervasive developmental disorders and mental retardation: diagnostic group differences, within-group distribution, and behavioral correlates. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(4), 491–499. https://doi.org/10.1097/00004583-200404000-00016

Murphy, V. E., Gibson, P. G., Giles, W. B., Zakar, T., Smith, R., Bisits, A. M., … Clifton, V. L. (2003). Maternal asthma is associated with reduced female fetal growth. *American Journal of Respiratory and Critical Care Medicine*, *168*(11), 1317–1323. https://doi.org/10.1164/rccm.200303-374OC

Murphy, V. E., Gibson, P., Talbot, P. I., & Clifton, V. L. (2005). Severe asthma exacerbations during pregnancy. *Obstetrics and Gynecology*, *106*(5 Pt 1), 1046–1054. https://doi.org/10.1097/01.AOG.0000185281.21716.02

Nagarajan, R. P., Hogart, A. R., Gwye, Y., Martin, M. R., & LaSalle, J. M. (2006). Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics*, *1*(4), e1-11.

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational Psychiatry*, *4*, e433. https://doi.org/10.1038/tp.2014.70

Ngo, S. T., Steyn, F. J., & McCombe, P. A. (2014). Gender differences in autoimmune disease. *Frontiers in Neuroendocrinology*, *35*(3), 347–369. https://doi.org/10.1016/j.yfrne.2014.04.004

Nguyen, A., Rauch, T. A., Pfeifer, G. P., & Hu, V. W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *24*(8), 3036–3051. https://doi.org/10.1096/fj.10-154484

Novakovic, B., Yuen, R. K., Gordon, L., Penaherrera, M. S., Sharkey, A., Moffett, A., … Saffery, R. (2011). Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. *BMC Genomics*, *12*, 529. https://doi.org/10.1186/1471-2164-12-529

Paquette, A. G., Houseman, E. A., Green, B. B., Lesseur, C., Armstrong, D. A., Lester, B., & Marsit, C. J. (2016). Regions of variable DNA methylation in human placenta associated with newborn neurobehavior. *Epigenetics*, *11*(8), 603–613. https://doi.org/10.1080/15592294.2016.1195534

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., … Clark, S. J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*, *17*(1), 208. https://doi.org/10.1186/s13059-016-1066-1

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., … Scherer, S. W. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, *94*(5), 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018

Price, M. E., Cotton, A. M., Lam, L. L., Farré, P., Emberly, E., Brown, C. J., … Kobor, M. S. (2013). Additional annotation enhances potential for biologically-relevant analysis of the

Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics & Chromatin*, *6*(1), 4. https://doi.org/10.1186/1756-8935-6-4

Raghavan, R., Riley, A. W., Volk, H., Caruso, D., Hironaka, L., Sices, L., … Wang, X. (2018). Maternal Multivitamin Intake, Plasma Folate and Vitamin B12 Levels and Autism Spectrum Disorder Risk in Offspring. *Paediatric and Perinatal Epidemiology*, *32*(1), 100–111. https://doi.org/10.1111/ppe.12414

Roberts, A. L., Lyall, K., Hart, J. E., Laden, F., Just, A. C., Bobb, J. F., … Weisskopf, M. G. (2013). Perinatal air pollutant exposures and autism spectrum disorder in the children of Nurses' Health Study II participants. *Environmental Health Perspectives*, *121*(8), 978–984. https://doi.org/10.1289/ehp.1206187

Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F., & Ronald, A. (2013). Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(13), 5258–5262. https://doi.org/10.1073/pnas.1211070110

Roifman, M., Choufani, S., Turinsky, A. L., Drewlo, S., Keating, S., Brudno, M., … Weksberg, R. (2016). Genome-wide placental DNA methylation analysis of severely growth-discordant monochorionic twins reveals novel epigenetic targets for intrauterine growth restriction. *Clinical Epigenetics*, *8*, 70. https://doi.org/10.1186/s13148-016-0238-x

Ronald, A., Happé, F., Dworzynski, K., Bolton, P., & Plomin, R. (2010). Exploring the relation between prenatal and neonatal complications and later autistic-like features in a representative community sample of twins. *Child Development*, *81*(1), 166–182. https://doi.org/10.1111/j.1467-8624.2009.01387.x

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., … State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA*, *311*(17), 1770–1777. https://doi.org/10.1001/jama.2014.4144

Schierding, W., Cutfield, W. S., & O'Sullivan, J. M. (2014). The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. *Frontiers in Genetics*, *5*, 39. https://doi.org/10.3389/fgene.2014.00039

Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–976. https://doi.org/10.1038/ng.940

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Schmidt, R. J., Hansen, R. L., Hartiala, J., Allayee, H., Schmidt, L. C., Tancredi, D. J., … Hertz-Picciotto, I. (2011). Prenatal vitamins, one-carbon metabolism gene variants, and risk for autism. *Epidemiology (Cambridge, Mass.)*, *22*(4), 476–485. https://doi.org/10.1097/EDE.0b013e31821d0e30

Schmidt, R. J., Tancredi, D. J., Ozonoff, S., Hansen, R. L., Hartiala, J., Allayee, H., … Hertz-Picciotto, I. (2012). Maternal periconceptional folic acid intake and risk of autism spectrum disorders and developmental delay in the CHARGE (CHildhood Autism Risks from Genetics and Environment) case-control study. *The American Journal of Clinical Nutrition*, *96*(1), 80–89. https://doi.org/10.3945/ajcn.110.004416

Schroeder, D. I., Blair, J. D., Lott, P., Yu, H. O. K., Hong, D., Crary, F., … LaSalle, J. M. (2013). The human placenta methylome. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 6037–6042. https://doi.org/10.1073/pnas.1215145110

Schroeder, D. I., Schmidt, R. J., Crary-Dooley, F. K., Walker, C. K., Ozonoff, S., Tancredi, D. J., … LaSalle, J. M. (2016). Placental methylome analysis from a prospective autism study. *Molecular Autism*, *7*, 51. https://doi.org/10.1186/s13229-016-0114-8

Shi, J., Marconett, C. N., Duan, J., Hyland, P. L., Li, P., Wang, Z., … Landi, M. T. (2014). Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nature Communications*, *5*, 3365. https://doi.org/10.1038/ncomms4365

Singer, A. B., Aylsworth, A. S., Cordero, C., Croen, L. A., DiGuiseppi, C., Fallin, M. D., … Daniels, J. L. (2017). Prenatal Alcohol Exposure in Relation to Autism Spectrum Disorder: Findings from the Study to Explore Early Development (SEED). *Paediatric and Perinatal Epidemiology*, *31*(6), 573–582. https://doi.org/10.1111/ppe.12404

Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.-Y., … Halperin, E. (2015). Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics & Chromatin*, *8*, 43. https://doi.org/10.1186/s13072-015-0035-3

Stark, M. J., Clifton, V. L., & Wright, I. M. R. (2009). Neonates born to mothers with preeclampsia exhibit sex-specific alterations in microvascular function. *Pediatric Research*, *65*(3), 292–295.

Straughen, J. K., Misra, D. P., Divine, G., Shah, R., Perez, G., VanHorn, S., … Salafia, C. M. (2017). The association between placental histopathology and autism spectrum disorder. *Placenta*, *57*, 183–188. https://doi.org/10.1016/j.placenta.2017.07.006

Suderman, M., Simpkin, A., Sharp, G. C., Gaunt, T., Lyttleton, O., McArdle, W., … Relton, C. L. (2017). Sex-associated autosomal DNA methylation differences are wide-spread and stable throughout childhood. *bioRxiv*, 118265. https://doi.org/10.1101/118265

Sun, W., Poschmann, J., Cruz-Herrera Del Rosario, R., Parikshak, N. N., Hajan, H. S., Kumar, V., … Prabhakar, S. (2016). Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell*, *167*(5), 1385–1397.e11. https://doi.org/10.1016/j.cell.2016.10.031

Surén, P., Roth, C., Bresnahan, M., Haugen, M., Hornig, M., Hirtz, D., … Stoltenberg, C. (2013). Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA*, *309*(6), 570–577. https://doi.org/10.1001/jama.2012.155925

Suter, M., Ma, J., Harris, A., Patterson, L., Brown, K. A., Shope, C., … Aagaard-Tillery, K. M. (2011). Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. *Epigenetics*, *6*(11), 1284–1294. https://doi.org/10.4161/epi.6.11.17819

Tarrade, A., Panchenko, P., Junien, C., & Gabory, A. (2015). Placental contribution to nutritional programming of health and diseases: epigenetics and sexual dimorphism. *The Journal of Experimental Biology*, *218*(Pt 1), 50–58. https://doi.org/10.1242/jeb.110320

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *57*(5), 585–595. https://doi.org/10.1111/jcpp.12499

Tran, P. L., Lehti, V., Lampi, K. M., Helenius, H., Suominen, A., Gissler, M., … Sourander, A. (2013). Smoking during pregnancy and risk of autism spectrum disorder in a Finnish National Birth Cohort. *Paediatric and Perinatal Epidemiology*, *27*(3), 266–274. https://doi.org/10.1111/ppe.12043

van Eijk, K. R., de Jong, S., Strengman, E., Buizer-Voskamp, J. E., Kahn, R. S., Boks, M. P., … Ophoff, R. A. (2015). Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood. *European Journal of Human Genetics: EJHG*, *23*(8), 1106–1110. https://doi.org/10.1038/ejhg.2014.245

Vatten, L. J., & Skjaerven, R. (2004). Offspring sex and pregnancy outcome by length of gestation. *Early Human Development*, *76*(1), 47–54.

Visser, J. C., Rommelse, N., Vink, L., Schrieken, M., Oosterling, I. J., van der Gaag, R. J., & Buitelaar, J. K. (2013). Narrowly versus broadly defined autism spectrum disorders: differences in pre- and perinatal risk factors. *Journal of Autism and Developmental Disorders*, *43*(7), 1505–1516. https://doi.org/10.1007/s10803-012-1678-6

Volk, H. E., Hertz-Picciotto, I., Delwiche, L., Lurmann, F., & McConnell, R. (2011). Residential proximity to freeways and autism in the CHARGE study. *Environmental Health Perspectives*, *119*(6), 873–877. https://doi.org/10.1289/ehp.1002835

Volk, H. E., Lurmann, F., Penfold, B., Hertz-Picciotto, I., & McConnell, R. (2013). Traffic-related air pollution, particulate matter, and autism. *JAMA Psychiatry*, *70*(1), 71–77. https://doi.org/10.1001/jamapsychiatry.2013.266

Wadhwa, P. D., Buss, C., Entringer, S., & Swanson, J. M. (2009). Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Seminars in Reproductive Medicine*, *27*(5), 358–368. https://doi.org/10.1055/s-0029-1237424

Walker, C. K., Anderson, K. W., Milano, K. M., Ye, S., Tancredi, D. J., Pessah, I. N., … Kliman, H. J. (2013). Trophoblast inclusions are significantly increased in the placentas of children in families at risk for autism. *Biological Psychiatry*, *74*(3), 204–211. https://doi.org/10.1016/j.biopsych.2013.03.006

Wang, M., Li, K., Zhao, D., & Li, L. (2017). The association between maternal use of folic acid supplements during pregnancy and risk of autism spectrum disorders in children: a meta-analysis. *Molecular Autism*, *8*, 51. https://doi.org/10.1186/s13229-017-0170-8

Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current Opinion in Neurology*, *26*(2), 146–153. https://doi.org/10.1097/WCO.0b013e32835ee548

Windham, G. C., Zhang, L., Gunier, R., Croen, L. A., & Grether, J. K. (2006). Autism spectrum disorders in relation to distribution of hazardous air pollutants in the san francisco bay area. *Environmental Health Perspectives*, *114*(9), 1438–1444.

Wong, C. C. Y., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., … Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry*, *19*(4), 495–503. https://doi.org/10.1038/mp.2013.41

Wu, S., Wu, F., Ding, Y., Hou, J., Bi, J., & Zhang, Z. (2017). Advanced parental age and autism risk in children: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *135*(1), 29–41. https://doi.org/10.1111/acps.12666

Xu, H., Wang, F., Liu, Y., Yu, Y., Gelernter, J., & Zhang, H. (2014). Sex-biased methylome and transcriptome in human prefrontal cortex. *Human Molecular Genetics*, *23*(5), 1260–1270. https://doi.org/10.1093/hmg/ddt516

Yong, W.-S., Hsu, F.-M., & Chen, P.-Y. (2016). Profiling genome-wide DNA methylation. *Epigenetics & Chromatin*, *9*, 26. https://doi.org/10.1186/s13072-016-0075-3

Yousefi, P., Huen, K., Davé, V., Barcellos, L., Eskenazi, B., & Holland, N. (2015). Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics*, *16*, 911. https://doi.org/10.1186/s12864-015-2034-y

Yuen, R. K., Peñaherrera, M. S., von Dadelszen, P., McFadden, D. E., & Robinson, W. P. (2010). DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. *European Journal of Human Genetics: EJHG*, *18*(9), 1006–1012. https://doi.org/10.1038/ejhg.2010.63

Zerbo, O., Qian, Y., Yoshida, C., Grether, J. K., Van de Water, J., & Croen, L. A. (2015). Maternal Infection During Pregnancy and Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *45*(12), 4015–4025. https://doi.org/10.1007/s10803-013-2016-3

Zhu, L., Wang, X., Li, X.-L., Towers, A., Cao, X., Wang, P., … Jiang, Y.-H. (2014). Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. *Human Molecular Genetics*, *23*(6), 1563–1578. https://doi.org/10.1093/hmg/ddt547

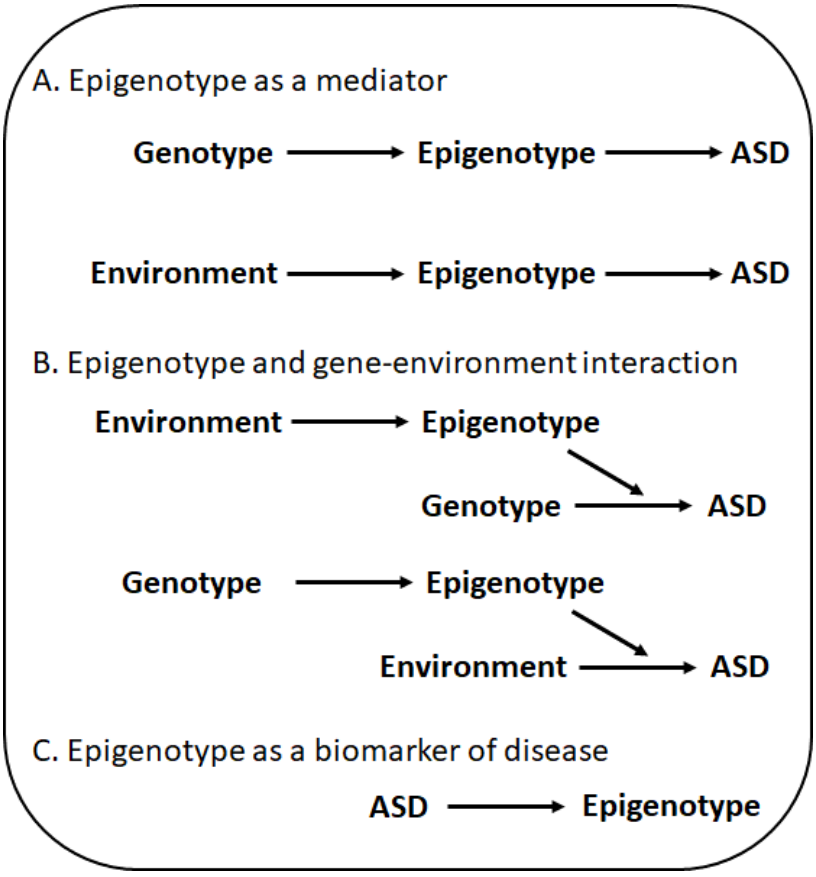



**Figure 1.1 Mechanisms for potential relationships of epigenotype and ASD.**

# CHAPTER 2: CROSS-TISSUE INTEGRATION OF GENETIC AND EPIGENETIC DATA OFFERS INSIGHT INTO AUTISM SPECTRUM DISORDER

Shan V. Andrews[1,2], Shannon E. Ellis[3], Kelly M. Bakulski[4], Brooke Sheppard[1,2], Lisa A. Croen[5], Irva Hertz-Picciotto[6,7], Craig J. Newschaffer[8,9], Andrew P. Feinberg[10,11], Dan E. Arking[2,3], Christine Ladd-Acosta[1,2,10*], M. Daniele Fallin[2,10,12*]

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St, Baltimore, MD 21205.

[2]Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St, Baltimore, MD 21205.

[3]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, 733 N. Broadway, Baltimore, MD 21205.

[4]Department of Epidemiology, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109.

[5]Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612.

[6]Department of Public Health Sciences, School of Medicine, University of California Davis, 4610 X St, Sacramento, CA 95817.

[7]MIND Institute, University of California Davis, 2825 50th St, Sacramento, CA 95817.

[8]AJ Drexel Autism Institute, Drexel University, 3020 Market St #560, Philadelphia, PA 19104.

[9]Department of Epidemiology and Biostatistics, Drexel University Dornsife School of Public Health, 3125 Market St, Philadelphia, PA 19104.

[10]Center for Epigenetics, Institute for Basic Biomedical Sciences, Johns Hopkins School of Medicine, 733 N. Broadway, Baltimore, MD 21205

[11]Department of Medicine, Johns Hopkins School of Medicine, 733 N. Broadway, Baltimore, MD 21205

[12]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Baltimore, MD, 21205.

## 2.1 Abstract

Integration of emerging epigenetic information with Autism Spectrum Disorder (ASD) genetic results may elucidate functional insights not possible via either type of information in isolation. Here, we use genotype and DNA methylation (DNAm) data from cord blood and peripheral blood to identify SNPs associated with DNA methylation (meQTL lists). Additionally, we use publicly available fetal brain and lung meQTL lists to assess enrichment of ASD GWAS results for tissue-specific meQTLs. ASD-associated SNPs are enriched for fetal brain (OR = 3.55; $P < 0.001$) and peripheral blood meQTLs (OR = 1.58; $P < 0.001$). The CpG targets of ASD meQTLs across cord, blood, and brain tissues are enriched for immune-related pathways, consistent with other expression and DNAm results in ASD, and reveal pathways not implicated by genetic findings. This joint analysis of genotype and DNAm demonstrates the potential of both brain and blood-based DNAm for insights into ASD and psychiatric phenotypes more broadly.

## 2.2 Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by deficits in social communication and interaction as well as restricted repetitive behavior(Rapin, 1997). ASD has a strong genetic basis(Colvert et al., 2015; Sandin et al., 2014), and most findings to date have been rare variants, including inherited and *de novo* mutations as well as copy number variations(Iossifov et al., 2014; O'Roak et al., 2014; Sanders et al., 2015). Although rare variants explain a relatively small proportion of all ASD cases(Gaugler et al., 2014), they provide converging evidence for three key biological processes implicated in ASD, including epigenetic regulation (Krumm, O'Roak, Shendure, & Eichler, 2014; Pinto et al., 2014; Sanders et al., 2015). Other lines of evidence also implicate epigenetic mechanisms in ASD(Abdolmaleky, Zhou, & Thiagalingam, 2015; Amir et al., 1999; Horsthemke & Wagstaff, 2008; Loke, Hannan, & Craig, 2015; Oberlé et al., 1991). Common genetic variation also plays a role, similar to other complex psychiatric diseases(Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), and mega-analysis GWAS results from the Psychiatric Genomics Consortium ASD workgroup (PGC-AUT)(Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017) are only recently available, and thorough examination of the biology implicated by common variants has not yet been fully pursued. Previous studies of neuropsychiatric disorders(Davis et al., 2012; Gamazon et al., 2013; van Eijk et al., 2015) have demonstrated the enrichment of GWAS results for expression quantitative trait loci (eQTLs), providing insights into the functional biology of discovered GWAS SNPs, assuming those SNPs

confer some risk through regulatory mechanisms. Given the implications of epigenetic regulation in ASD, a similar approach exploring enrichment of SNPs controlling epigenetic marks, such as DNA methylation, may be fruitful, assuming similarly that ASD genetic risk may act in part through epigenetic regulation. As with expression loci, genetic variation contributes to DNAm levels locally and distally(Hannon et al., 2016; Shi et al., 2014) and thus integration of methylation quantitative trait loci (meQTLs), or SNPs that are highly associated with DNAm, and autism-associated GWAS results may inform our understanding of autism GWAS findings. Moreover, meQTLs are enriched in top hits for bipolar disorder(Gamazon et al., 2013) and schizophrenia(Hannon et al., 2016; van Eijk et al., 2015), which have well-established genetic overlap with ASD(Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013).

Epigenetic patterns necessarily differ across tissue types, given their role in cell differentiation and expression. For brain-related conditions like ASD, careful consideration of tissue source for epigenetic analyses is warranted(Bakulski, Halladay, Hu, Mill, & Fallin, 2016). This must be balanced by consideration of tissue availability, which is limited for ASD brain tissue. We and others have shown blood-based epigenetic biomarkers are useful in psychiatric conditions, including ASD(Montano et al., 2016; Wong et al., 2014), while recognizing the limitations and need for comparison to brain-derived data wherever possible(Bakulski et al., 2016; Davies et al., 2012; Hannon, Lunnon, Schalkwyk, & Mill, 2015). ASD-related epigenetic differences have also been observed in buccal(Berko et al., 2014), lymphoblastoid cell line(Nguyen, Rauch, Pfeifer, & Hu, 2010), and postmortem brain samples(James, Shpyleva, Melnyk, Pavliv, & Pogribny, 2014; Ladd-Acosta et al., 2014; Nardone et al., 2014), as well as in the sperm from fathers of children with ASD(Feinberg et al., 2015). Blood-brain DNAm concordance studies have not frequently

observed high correlation of DNAm levels at specific sites across tissues; however, when such concordance is observed, it is likely due to genetic influences(Davies et al., 2012; Hannon et al., 2015). meQTL signals overlap in adult brain and blood tissues(Smith et al., 2014), suggesting blood-derived meQTLs may also reflect SNP-DNAm relationships in brain tissue, though this relationship has rarely been tested.

This study used lists and locations of meQTLs and their CpG targets, termed meQTL maps, from cord blood, peripheral blood, and fetal brain tissues to characterize and prioritize ASD GWAS SNPs and the CpG sites under their control. We created such meQTL maps from our own data for infant cord and childhood peripheral blood tissues, and used publicly available meQTL maps for brain and lung tissues to examine cross-tissue meQTLs. We find that ASD GWAS signals are enriched for meQTLs in peripheral blood and fetal brain. The CpG site targets controlled by ASD-associated SNPs are enriched for immune response pathways, and can implicate new genes not directly identified by GWAS results alone. Finally, we extend the characterization of SNP-controlled CpG sites to neuropsychiatric disease more generally, and discover their enrichment for specific regulatory features within and across tissue type. Our work demonstrates the utility of jointly analyzing GWAS and DNAm data for insights into ASD and neuropsychiatric disease.

## 2.3 Methods

### *2.3.1 Cord blood samples*

Cord blood DNA was obtained from newborn participants of the Early Autism Risk Longitudinal Risk Investigation (EARLI), an enriched-risk prospective birth cohort described in detail elsewhere(Newschaffer et al., 2012). The EARLI study was approved by Human Subjects Institutional Review Boards (IRBs) from each of the four study sites (Johns Hopkins University,

Drexel University, University of California Davis, and Kaiser Permanente). Informed consent was obtained from all participating families. The 232 mothers with a subsequent child born through this study had births between November 2009 and March 2012. Infants were followed with extensive neurophenotyping until age three, including ASD diagnostics.

*2.3.2 Cord blood DNA methylation*

Cord blood biospecimens were collected and archived at 175 births. DNA was extracted using the DNA Midi kit (Qiagen, Valencia, CA) and samples were bisulfite treated and cleaned using the EZ DNA methylation gold kit (Zymo Research, Irvine, CA). DNA was plated randomly and assayed on the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA), or "450k", at the Center for Inherited Disease Research (CIDR, Johns Hopkins University). Methylation control gradients and between-plate repeated tissue controls (n=68) were used(Feinberg et al., 2015).

The *minfi* library (version 1.12)(Aryee et al., 2014) in R (version 3.1) was used to process raw Illumina image files with the background correcting and dye-bias equalization method: normal-exponential using out-of-band probe (Noob)(Fortin et al., 2014; Triche, Weisenberger, Van Den Berg, Laird, & Siegmund, 2013). Probes with failed detection *P*-value (>0.01) in >10% of samples were removed (n=661), as were probes annotated as cross-reactive (n=29,233)(Chen et al., 2013) and those mapping to sex chromosomes (n=11,648). All cord samples passed sample-based filters (sex matching, detection *P*-values > 0.01 in greater than 1% of sites). Pre-processed data were adjusted for batch effects related to hybridization date and array position using the *ComBat()* function(W. E. Johnson, Li, & Rabinovic, 2007) in the *sva* R package (version

3.9.1)("sva," n.d.). Methylation data were available from 175 cord blood samples at 445,241 probes.

*2.3.3 Cord blood genotyping*

Overlapping cord blood DNA methylation and corresponding SNP data was available on 171 EARLI cord blood samples. Genotype data were generated for 841 EARLI family biosamples and 18 HapMap control samples run on the Omni5 plus exome (Illumina, San Diego, CA) genotyping array at CIDR (Johns Hopkins University), generating data on 4,641,218 SNPs. The duplicated HapMap sample concordance rate was 99.72% and the concordance rate among five EARLI samples with blind duplicates was 99.9%. Samples were removed if they were HapMap controls (n=18), technical duplicates (n=5; selected by frequency of missing genotypes), or re-enrolled families/other relatedness errors (n=9). No samples met the following additional criteria for exclusion: missing genotypes at >3% of probes, or excess heterozygosity or homozygosity (4 SD). Probes were removed for CIDR technical problems (n=94,712), missing genomic location information (n=8,124). Among probes with high minor allele frequencies (>5%), SNPs with a missing rate $\geq$ 5% were excluded (n=8,902) and among probes with low minor allele frequencies (<5%) SNPs with a missing rate >1% were excluded (n=65,855). There were 827 samples and 4,463,625 probes at this stage and SNPs out of Hardy-Weinberg equilibrium ($\chi^2$ test $P< 10\text{-}7$) were flagged (n=2,170). Samples were merged with the 1,000 genomes project (1000GP, version 5) data(1000 Genomes Project Consortium et al., 2012) and EARLI ancestries were projected into four categories (White, Black, Asian, Hispanic). EARLI measured genotype data was phased using SHAPEIT(Delaneau, Zagury, & Marchini, 2013) and imputed to the 1000GP data using Minimac3(Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012). SNPs with MAF > 1% were retained, leaving a total of 9,377,008 SNPs.

*2.3.4 Peripheral blood samples*

Samples were obtained from the Study to Explore Early Development (SEED), a multi-site, national case-control study of children aged 3-5 years with and without ASD. Overall, 2,800 families were recruited and classified into 3 groups according to the status of the child: the ASD group, the general population control group, and the (non-ASD) developmental delay group(Schendel et al., 2012). This study was approved as an exemption from the Johns Hopkins Institutional Review Board (IRB) under approval 00000011. Informed consent was obtained from all participants as part of the parent SEED study. SEED recruitment was approved by the IRBs of each recruitment site: Institutional Review Board (IRB)-C, CDC Human Research Protection Office; Kaiser Foundation Research Institute (KFRI) Kaiser Permanente Northern California IRB, Colorado Multiple IRB, Emory University IRB, Georgia Department of Public Health IRB, Maryland Department of Health and Mental Hygiene IRB, Johns Hopkins Bloomberg School of Public Health Review Board, University of North Carolina IRB and Office of Human Research Ethics, IRB of The Children's Hospital of Philadelphia, and IRB of the University of Pennsylvania. All enrolled families provided written consent for participation.

*2.3.5 Peripheral blood DNA methylation*

Genomic DNA was isolated from whole blood samples using the QIAsumphonia midi kit (Qiagen, Valencia, CA). For each a subset of case and control samples (n = 630), bisulfite treatment was performed using the 96-well EZ DNA methylation kit (Zymo Research, Irvine, CA). Samples were randomized within and across plates to minimize batch and position effects. The minfi R package (version 1.16.1) was used to process Illumina .idat files generated from the array(Aryee et al., 2014). Control samples (n=14) were removed and quantile normalization performed using the *minfi* function *preprocessQuantile()* (Touleimat & Tost, 2012).Probes with

failed detection *P*-value (>0.01) in >10% of samples were removed (n=772), as were probes annotated as cross-reactive (n=29,233)(Chen et al., 2013), and probes on sex chromosomes (n=11,648). Samples were excluded if reported sex did not match predicted sex (*minfi* function *getSex()*) (n=0), detection *P*-values > 0.01 in greater than 1% of sites (n=2), low overall intensity (median methylated or unmethylated intensity < 11; n=2), and if they were duplicates (n=8). Successive filtering according to these criteria resulted in 445,154 probes and 604 samples.

*2.3.6 Peripheral blood genotyping*

Of the SEED samples with DNAm data, 590 had whole-genome genotyping data available, measured using the Illumina HumanOmni1-Quad BeadChip (Illumina, San Diego, CA). Standard quality control measures were applied: removing samples with < 95% SNP call rate, sex discrepancies, relatedness (Pi-hat > 0.2), or excess hetero- or homozygosity; removing markers with < 98.5% call rate, or monomorphic. Phasing was performed using SHAPEIT(Delaneau et al., 2013) followed by SNP imputation via the IMPUTE2 software(Howie et al., 2012), with all individuals in the 1000 Genomes Project as a reference. Genetic ancestry was determined using EigenStrat program(Price et al., 2006). A total of 4,948,723 SNPs were available post imputation at MAF > 1%.

*2.3.7 Normal lung tissue meQTLs*

A list of meQTLs identified in a recent characterization of normal lung tissue(Shi et al., 2014) as well as the total list of SNPs (n = 569,753) and 450k CpG sites (n = 338,730) tested for meQTL identification (i.e. passed filtering and QC done in that study) was obtained from the study authors.

*2.3.8 Fetal brain meQTLs*

Fetal brain meQTLs were identified via imputed genotypes in a recent study examining meQTLs in the context of schizophrenia(Hannon et al., 2016). The total list of SNPs (n = 5,159,699) and 450k CpG sites (n = 314,554) that were tested (i.e. passed filtering and QC done in that study) was obtained from the study authors. For all analyses, only fetal brain meQTLs within a SNP to CpG distance of 1 Mb, were included, in order to improve comparability to the other 3 meQTL lists, where distant (trans) meQTL relationships were not explored (peripheral blood, cord blood) or used (lung).

*2.3.9 meQTL identification parameters*

There are three main parameters of interest in a meQTL query: the SNP minor allele frequency (MAF) threshold for inclusion, the definition of standard deviation cutoff that dictates a CpG site is variably methylated, and the maximum physical distance between a SNP and CpG site to be queried, often referred to as the window size. These 3 factors contribute to the total number of CpG to SNP linear regression tests that are performed. Our available sample sizes (and thus statistical power, at fixed effect size) for the joint DNAm and genotype data differed for peripheral blood (n = 339) and cord blood (n = 121) analyses after limiting both to samples of European descent identified via principle components analysis of SNP data. Thus, the ideal combination of these parameters should differ between the two study populations for them to be comparable.

For each tissue sample set, we computed the total number of CpG to SNP linear regression tests at various combinations of the 3 main parameters of interest to a meQTL query. We then used the genetic power calculator Quanto(Gauderman, 2002) to determine the most permissive set of parameters that allowed for 80% power to detect a 5% difference in methylation for each

addition of the minor allele, at the lowest allowed MAF. We computed this power calculation at a Bonferroni-based significance level derived from the total number of CpG to SNP linear regression tests. We defined 'most permissive' in a hierarchical manner that first prioritized the inclusion of the most methylation sites (lowest sd cutoff), then the inclusion of the most number of SNPs (lowest MAF threshold), and then the use of the largest window size. This procedure resulted in study-specific MAF thresholds for the SNP data, standard deviation cutoffs for the methylation data, and window sizes that were tailored to the number of samples available.

*2.3.10 meQTL identification procedure*

Pairwise associations between each SNP and CpG site were estimated via the R package *MatrixEQTL*(Shabalin, 2012), with percent methylation (termed 'Beta value', ranging from 0 to 100) regressed onto genotype assuming an additive model, adjusting for the first two principal components of ancestry and sex. Models did not adjust for age given the very narrow age ranges in each tissue type.

False discovery rate (FDR) was controlled via permutation(Shi et al., 2014). Briefly, the total number of CpG sites ($N_{obs}$) under genetic control was obtained for a meQTL *P*-value of $p_o$. Genome-wide meQTL query was performed for each of 100 permuted sets of the genotype data (scrambling sample IDs, to retain genotype correlation structure). In each set, we retained the total number of CpG sites under genetic control ($N_{null}$) at the same *P*-value $p_o$. The FDR was defined as the mean($N_{null}$)/ $N_{obs}$. Finally we determined the value of $p_o$ to control the FDR at values of 10%, 5%, and 1%. Both the meQTL discovery and FDR determination were performed in each tissue or study sample.

*2.3.11 Enrichment of meQTLs in ASD-associated SNPs*

We tested for enrichment of meQTLs from four tissue types among ASD GWAS SNPs. ASD SNPs were assigned from the PGC-AUT analysis (downloaded February 2016), based on 5,305 cases and 5,305 pseudocontrols("Psychiatric Genomics Consortium," n.d.). The PGC provides results for 9,499,589 SNPs; 11,749 SNPs exceeded an ASD *P*-value threshold of 1e-03, and 1,094 SNPs exceeded an ASD *P*-value threshold of 1e-04. For each tissue, we included only SNPs available in both the PGC-AUT analysis and our meQTL analysis, either via direct or proxy ($r^2 > 0.8$ within 500Kb window in CEU 1KG) overlap as defined via the SNAP software(A. D. Johnson et al., 2008).

To estimate the proportion of meQTLs among ASD SNPs versus among all SNPs (or a sample of null SNPs), we recognized three important factors that could differ between null SNP sets and the ASD SNP set: LD structure, MAF distribution, and number of CpG sites per window size in the meQTL screen. We designed a comparison process to address each of these. First, we performed LD pruning 'supervised' by PGC ASD *P*-value (so as to not prune away all ASD SNPs) using PriorityPruner (v0.1.2)("PriorityPruner," n.d.), removing SNPs at $r^2 > 0.7$ within a sliding 500Kb window. For the peripheral blood and cord blood datasets this pruning was done with the study-specific genotype data, and for the fetal brain and lung datasets this pruning was done with 1000 Genomes CEU samples. Second, we grouped remaining SNPs into MAF bins of 5%. Third, we characterized each SNP according to the number of CpGs within the meQTL discovery window size to allow for differential opportunity to have been identified as a meQTL. We then collapsed this number into categories of 0-49, 50-99, etc. to reflect the same concept. We defined 1,000 null SNP sets by finding, for each SNP in the ASD set, a random SNP in the genome that matched that SNP on both MAF bin and CpG opportunity. We computed an

enrichment fold statistic as the proportion of meQTLs in the ASD SNP set divided by the mean proportion of meQTLs across null sets; and a *P*-value as the total number of null set proportions as or more extreme than in the ASD set. To evaluate the robustness of our results, we used two PGC AUT *P*-value cutoffs (1e-03, 1e-04) and three meQTL *P*-value cutoffs (FDR 10%, 5%, 1%) for peripheral blood and cord blood. However, based on available information for lung and fetal brain, we were limited to assess our results at FDR 5% for lung, and *P*<1e-08 for fetal brain for meQTL *P*-value.

*2.3.12 Gene ontology analysis of meQTL targets*

We further assessed which biological pathways are implicated by the location of CpG targets of ASD SNPs acting as meQTLs. To do this, we identified CpG sites associated with ASD SNPs (ASD-related meQTL targets) among all CpG sites controlled by SNPs (all meQTL targets). We then examined Gene Ontology (GO) terms specific to these ASD-related meQTL targets, in order to enumerate biological pathways engaged specifically by ASD SNPs.

Specifically, we first filtered the full list of CpG sites associated with any meQTL to only those sites associated with an ASD SNP (PGC Wald $P < 1e\text{-}04$; N=1,094) or their proxies ($r^2 > 0.8$ within 500kb window in CEU 1KG as defined via the SNAP software(A. D. Johnson et al., 2008)). We used thresholds of FDR ≤ 5% for peripheral and cord blood meQTL lists, and Wald $P < 1e\text{-}08$ for the fetal brain list. We only examined CpG sites that did not overlap with SNPs within 10bp of the CpG site or at the single base extension(Bibikova et al., 2011), as it has been previously demonstrated that these CpG sites may strongly influence functional-type enrichment analysis of CpG sites(McClay et al., 2015), and these CpG sites were not examined in the fetal brain meQTL lists(Hannon et al., 2016). We used the *gometh()* function in the *MissMethyl* R

package("missMethyl," n.d.), which maps 450k DNAm sites to their nearest gene, and corrects for bias due to non-uniform coverage of genes on the 450k. We further ran nominally significant (hypergeometric test $P< 0.05$) results for the category "biological processes" through the REVIGO tool to avoid reporting GO terms with a greater than 70% overlap in gene lists(Supek, Bošnjak, Škunca, & Šmuc, 2011). Finally, we determined the set of terms in these lists that overlapped at least two tissues, and prioritized them by summing the scaled, enrichment *P*-value-based rank in each tissue. This scaling was done by dividing the raw rank for the term in the list for that tissue by the total number of nominally significant, post-REVIGO terms for that tissue.

We also ran analogous GO analyses comparing all meQTL targets to all CpGs to explore functional implications for meQTL targets versus CpGs not under strong genetic control. This allowed for comparison of ASD SNP-specific functional pathways engaged through methylation versus general SNP functional pathways engaged through methylation.

*2.3.13 Identifying novel genes via ASD meQTL target locations*

Defining ASD SNPs as those with PGC Wald $P < 1e-04$, meQTL relationships as in the GO analysis, and RefSeq genes from the UCSC Genome Browser(Karolchik et al., 2004), we annotated gene overlap (if any) via *findOverlaps()* in the *GenomicRanges* R package for all ASD SNPs and their associated CpG sites (if any). We filtered out long intergenic non-coding RNAs, long non-coding RNAs, microRNAs, and small associated RNAs from the RefSeq gene list. We further collapsed SNPs into bins by LD block. Blocks were defined using recombination hot spot data from 1000 Genomes(1000 Genomes Project Consortium et al., 2012).

*2.3.14 Regulatory feature characterization of meQTL targets*

To quantify the propensity of regulatory features to overlap with meQTL targets within and across tissue type, we first compared regulatory feature overlap of all meQTL targets to non-meQTL targets. We next compared meQTL targets of psychiatric condition-related SNPs to meQTL targets of SNPs unrelated to psychiatric conditions. SNPs associated with psychiatric conditions were obtained from the PGC cross-disorder analysis(Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013) (PGC Wald $P < 1e-04$) and their proxies. We used these SNPs in order to analyze a greater total number of meQTL targets than associated with ASD SNPs only (and thus ensure a well-powered analysis), and to make functional insights that could be applied to psychiatric disease more broadly. Our results are still relevant to ASD in light of known cross-disorder GWAS consistency(Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013).

We performed both comparisons for unique and overlapping tissue categories (n =7): peripheral blood, cord blood, fetal brain, lung, intersection of peripheral blood and cord blood, intersection of peripheral blood and fetal brain, and intersection of peripheral blood and lung. For each intersection, we conducted a new meQTL discovery screen in which the peripheral blood was down sampled to the sample size of the other tissue, and run at the same parameters used to identify meQTLs in that tissue. This increases comparability with respect to power and meQTL query parameters. For the peripheral blood overlaps with cord blood and lung, we also computed the meQTL *P*-value to control the FDR at 5%, since this FDR threshold was available for each of those tissue types(Shi et al., 2014). However, we only computed the FDR *P*-values using data from the first 6 chromosomes, as we found empirically that FDR *P*-value estimates stabilized by this point. Finally, for the peripheral blood-fetal brain comparison, we retained results for

peripheral blood that passed a meQTL *P*-value of 1E-8, as reported from the fetal brain study(Hannon et al., 2016).

Regulatory feature information came from several sources. General DHSs were defined as those CpG probes experimentally determined to be within a DHS, as determined by the manifest for the 450k array("IlluminaHumanMethylation450kmanifest," n.d.). In addition, tissue-specific DHS data were tested for enrichment. Brain DHSs were downloaded from GEO(Edgar, Domrachev, & Lash, 2002) for three brain regions: Frontal Cortex [GEO Sample ID: GSM1008566], Cerebellum [GSM1008583], and Cerebrum [GSM1008578]). Two blood (CD14+ Monocytes; 'wgEncodeOpenChromDnaseMonocd14' and CD4+ cells; 'wgEncodeUwDnaseCd4naivewb78495824PkRep1') and one lung-derived (IMR90; 'wgEncodeOpenChromDnaseImr90Pk') data sets were additionally downloaded from the UCSC Genome Browser(Karolchik et al., 2004).

Tissue-specific histone data were compiled from the Roadmap Epigenomics Project(Roadmap Epigenomics Consortium et al., 2015) for five different marks: H3K27me3, H3K36me3, H3K4me1, H3K4me3, and H3K9me3. As Epigenome Roadmap Project data were often generated across a number of individuals, for those cases in which data were generated in more than one Caucasian individual, the overlap across individual samples was utilized in downstream analyses. Overlap was calculated using the UCSC Genome Browser's '*intersect*' function for those samples indicated in **Table 2.4**. Regions with any overlap were included in functional enrichment analyses.

Finally, TFBS information from ChIP-Seq experiments carried out by the ENCODE project(ENCODE Project Consortium, 2012) were extracted for 161 transcription factors from the UCSC Genome Browser ('wgEncodeRegTfbsClusteredV3')(Karolchik et al., 2004).

Significant feature overlap was assessed via two-sided Fisher's 2x2 exact test, with Bonferroni correction ($P < 0.05 / (181 \text{ regulatory features} * 7 \text{ categories}) = 3.95e\text{-}05$). Odds ratio and *P*-value were recorded for each test in each unique and overlapping tissue category.

## 2.4 Results

### *2.4.1 Creating meQTL maps*

We identified meQTL SNPs using combined GWAS and 450K methylation array data available on both peripheral blood and cord blood samples. For these analyses, we defined study-specific parameters that were optimal for each dataset and determined the *P*-value (Wald test) to control the FDR at 10%, 5%, and 1%. In peripheral blood, we identified 1,878,577 meQTLs controlling DNAm at 85,250 CpGs; in cord blood, we found 1,252,498 meQTLs controlling DNAm at 35,905 CpGs, both at FDR = 5%. Peripheral blood and cord blood meQTLs, on average, were associated with 4.83 and 2.56 CpG sites respectively. Statistical significance was inversely related to distance between SNP and CpG site (**Supplementary Figure I-1**). We have provided a full list of all identified peripheral and cord blood meQTLs and their associated CpG sites at FDR = 5% (**Supplementary Data II-1 and II-2**).

We used publicly available lung(Shi et al., 2014) (to include a likely non ASD-related tissue) and fetal brain(Hannon et al., 2016) meQTL lists and the *P*-value cutoffs stated in those respective studies (Wald $P = 1e\text{-}08$ for fetal brain and Wald $P = 4e\text{-}05$ = FDR 5% for lung). In fetal brain,

there were a total of 299,992 meQTLs controlling 7,863 CpGs, and in lung there were 22,866 meQTLs controlling 34,304 CpG sites. Dataset characteristics, meQTL parameters, and *P*-values used are summarized in **Table 2.1**. In all tissues, meQTL targets (CpG sites controlled by meQTLs) implicate additional genes that are not accounted for by their corresponding meQTLs (**Supplementary Table I-1**).

There were 2,704,013 overlapping SNPs considered for meQTL discovery across peripheral blood, cord blood, and fetal brain analyses. Of these, 125,869 (4.65%) were identified as meQTLs in all three tissues, 407,722 (15.08%) were meQTLs only in peripheral and cord blood, 30,691 (1.14%) were meQTLs only in peripheral blood and fetal brain, and 528 (0.02%) were meQTLs only in cord blood and fetal brain (**Supplementary Table I-2**).

*2.4.2 Enrichment of meQTLs in ASD GWAS SNPs across 4 tissue types*

We observed enrichment of fetal brain meQTLs at both the more liberal GWAS SNP *P*-value threshold (enrichment fold = 1.70, permutation $P_{enrichment}$ < 1e-03 at Wald $P_{GWAS}$ < 1e-03), and at a more stringent GWAS *P*-value threshold (3.55, permutation $P_{enrichment}$ < 1e-03 at Wald $P_{GWAS}$ < 1e-04) (**Table 2.2**). There was no association with lung meQTLs at either the more liberal (1.09, permutation $P_{enrichment}$ = 0.343) or more stringent (0.80, permutation $P_{enrichment}$ = 0.301) threshold.

In peripheral and cord blood, we considered multiple GWAS SNP *P*-value thresholds as well as multiple meQTL discovery thresholds (the latter not available in brain and lung public data). There was significant meQTL enrichment for all GWAS and meQTL thresholds considered using peripheral blood meQTLs (enrichment fold range = 1.20 – 1.58, permutation $P_{enrichment}$ <

1e-03; **Table 2.2**). However, in cord blood, meQTL enrichment was only observed for a liberal GWAS SNP threshold (range = 1.14 – 1.21, permutation $P_{enrichment}$ = 0.011 – 0.032 at Wald $P_{GWAS}$ < 1e-03). This was not statistically significant after considering a Bonferonni correction to account for the 16 enrichment tests performed.

*2.4.3 Gene Ontology enrichment analyses of meQTL targets*

We examined the biological functions of meQTL targets of ASD SNPs specifically compared to meQTL targets generally. We identified 210, 66, and 53 meQTL targets associated with ASD SNPs in peripheral blood, cord blood, and fetal brain respectively. After mapping these CpG sites to genes, performing GO enrichment analyses, and removing overlapping GO terms, there were a total 95, 76, and 47 nominally significant (hypergeometric test $P < 0.05$) biological processes, respectively.

A total of 37 biological processes were present across either two or three tissues (**Table 2.3, Supplementary Data II-3 to II-5**), many of them relating to immune system function. Of these, 3 terms overlapped across all three tissues, 12 processes were enriched in cord blood and fetal brain but not peripheral blood, and 22 processes were present in both the peripheral and cord blood but not in fetal brain.

To test whether our findings were unique to ASD meQTL targets, we performed the same analysis comparing all meQTL targets to all CpG sites. (**Supplementary Figures I-2 to I-4**). Though there were some immune-related pathways discovered for fetal brain ASD meQTL targets that are also enriched in meQTLs generally, this was not the case in peripheral and cord blood.

*2.4.4 Expansion of ASD GWAS loci via ASD meQTL target locations*

The location of CpG targets for particular meQTL associations can further elucidate genes or regions relevant to ASD risk beyond the genomic location of the associated SNP variant. Of the 1,094 ASD-associated PGC SNPs (Wald $P < 1e\text{-}04$), five (0.46%) were detected as meQTLs across peripheral blood, cord blood, and fetal brain tissues (**Supplementary Table I-3, Supplementary Data II-6**). Consideration of the CpG DNAm targets of these SNPs implicates genes not directly annotated to the SNPs themselves. For example, ASD SNPs in *XKR6* target CpGs in *TDH* in both peripheral blood and fetal brain, and target CpGs in *SOX7* in peripheral blood and cord blood (**Figure 2.1A**). A similar result can be seen for ASD SNPs in *PPFIA3* with meQTL target CpGs that implicate *HRC* (**Figure 2.1B**).

*2.4.5 Characterizing meQTL targets for regulatory feature overlap*

We sought to quantify the propensity of regulatory features to overlap with meQTL targets within and across tissue type, and particularly whether meQTL targets of SNPs associated with psychiatric conditions have specific regulatory features. Individual and overlapping tissue meQTL target lists were compared for regulatory feature annotation. First, among psychiatric condition associated SNPs (via the PGC cross-disorder analysis(Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013)), their meQTL targets were significantly enriched for DNaseI hypersensitive sites (DHSs) in peripheral blood (OR = 1.22, Fisher's exact $P = 0.014$), fetal brain (OR = 2.23, Fisher's exact $P = 3.5e\text{-}03$), and peripheral blood-fetal brain overlap lists (OR = 2.22, Fisher's exact $P = 0.018$; black font and boxes, **Figure 2.2**), compared to meQTL targets of SNPs not associated with psychiatric conditions. Further, there was marginally significant enrichment of CD14 cell-specific DHSs (OR = 2.42, Fisher's exact $P = 0.013$;

**Supplementary Data II-7**) in the peripheral blood-fetal brain list. Few chromatin marks met Bonferroni significance (Fisher's exact $P \leq 3.95e\text{-}05$) defined by the 181 tests of regulatory features performed in all 7 lists of meQTL targets, though numerous marginally significant enrichment associations were observed for blood H3K36me3 (active) and blood H3K27me3 (repressive). Transcription factor binding sites (TFBSs) with observed enrichment include (**Supplementary Data II-7**) STAT1 for fetal brain (OR = 4.32, Fisher's exact $P = 2.66e\text{-}05$) and peripheral blood (OR = 2.24, Fisher's exact $P = 3.56e\text{-}08$), TAF1 for peripheral blood (OR = 1.53, Fisher's exact $P = 2.24e\text{-}06$), cord blood (OR = 2.24, Fisher's exact $P = 4.01e\text{-}06$), and fetal brain (OR = 3.2, Fisher's exact $P = 4.40e\text{-}06$), and POL2RA for peripheral blood (OR = 1.38, Fisher's exact $P = 1.14e\text{-}06$), cord blood (OR = 2.28, Fisher's exact $P = 3.54e\text{-}08$), and their overlap (OR = 2.20, Fisher's exact $P = 9.63e\text{-}09$).

When considering meQTL targets generally (regardless of their status of being downstream of PGC SNPs), compared to non-meQTL-target CpGs, enrichment was observed for DHSs for all 7 meQTL target lists, with the largest effect sizes among the peripheral blood-cord blood overlap list and the peripheral blood-fetal brain overlap list (gray font and boxes, **Figure 2.2**). In fact, a large number of regulatory features were significantly enriched among meQTL targets for each tissue, including lung, given the very large sample size of CpG sites (**Supplementary Data II-8**). However, these were typically small effects. Larger, but still moderate enrichment effects were primarily seen for cross-tissue overlap meQTL lists, particularly for the peripheral blood-cord blood overlap list and the peripheral blood-fetal brain overlap list.

## 2.5 Discussion

This first study integrating ASD GWAS results and meQTL maps provides novel insights about ASD etiology using data within and across tissue types. First, using blood samples from birth and early life, we identify meQTL maps and compare them to previously reported fetal brain tissue meQTLs, showing a subset of SNPs that are meQTLs across all three tissues. , The highest percent overlap is between peripheral and cord blood meQTL maps, as is expected given their tissue similarity. When examining enrichment among ASD GWAS results, we observe enrichment of peripheral blood ($1.20 \leq OR \leq 1.58$; permutation $P < 0.001$) and fetal brain (OR = 1.70 and 3.55; permutation $P < 0.001$) meQTLs. When considering the biological processes annotated to ASD meQTL targets, we see enrichment for immune-related pathways using all three tissue meQTL maps. Further, specific ASD meQTL targets may suggest novel regions for functional follow-up of ASD genetic associations. Finally, we identify several regulatory elements that preferentially overlap with meQTL targets associated with known SNPs for neuropsychiatric disease generally. Our results demonstrate the utility of meQTLs and their CpG targets for insights into ASD and neuropsychiatric disease overall.

Comparison of meQTL lists across tissues presents several challenges for interpretation of results. First, each set of samples came from a different study source, reflecting different sets of individuals and different sampling strategies, as well as differences in sample size and in genotyping and methylation array platforms. For example, we expected and observed considerable overlap between cord and peripheral blood meQTL signals, and less overlap with brain. The lack of further cross-tissue concordance with brain and blood could be due to differential statistical power between studies, lack of SNP or CpG overlap on arrays, or differences in pipelines used for meQTL discovery (choices of window size, SNP minor allele

frequency, etc.). In our functional characterization of meQTL targets, we used down sampling of peripheral blood results to make them comparable to the sample size and meQTL query pipeline decisions of the tissue to which it was being compared. While this is likely an incomplete solution, it is a step toward harmonization that has not been carried out in other studies.

We demonstrate that joint analysis of SNP and DNAm data can reveal novel insights towards ASD etiology not apparent when looking at either type of data alone. It is important to examine the biological implications of the genes implicated by SNPs, as well as the genes and regulatory functions implicated by DNAm. When considering the ASD SNPs, we found enrichment of fetal brain and peripheral blood meQTLs. The enrichment was stronger at increasingly stringent meQTL *P*-value and ASD *P*-value thresholds, bolstering confidence in these findings. These results are also concordant with similar studies of schizophrenia, a disorder with known genetic overlap to ASD(Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013), that have demonstrated enrichment in fetal brain meQTLs(Hannon et al., 2016) and peripheral blood meQTLs(van Eijk et al., 2015). A previous study examining enrichment of eQTLs in ASD GWAS SNPs observed enrichment in parietal and cerebellar eQTLs but not lymphoblastoid cell line eQTLs(Davis et al., 2012), though the GWAS results in that report likely differ greatly from those of the larger PGC-AUT mega-analysis. Crucially, we did not observe enrichment of lung meQTLs, supporting the specificity of fetal brain and peripheral blood results. However, we also did not observe an enrichment of cord blood meQTLs, suggesting the role of ASD-related DNAm marks in peripheral tissues may be developmentally regulated or a function of age.

Additional insights may be gained through examination of specific CpG targets of the ASD-related SNPs. Among CpG sites that are targets of ASD SNP meQTLs, there is an abundance of

immune response-related pathways, using brain, peripheral blood, or cord blood meQTL lists. This immune enrichment was not seen when considering CpG targets of all meQTLs in blood (not just the ASD SNPs), suggesting specificity to ASD. However, such enrichment was seen for all meQTL targets in fetal brain. This may be a consequence of the complications during pregnancy that resulted in fetal tissue collection (56-166 days post conception(Hannon et al., 2016)). Though many immune-related disorders are known to be comorbid with ASD(Estes & McAllister, 2015), previous enrichment-type analysis for genetic variants alone have not highlighted immune-related pathways, instead implicating chromatin regulation, synaptic function, and Wnt signaling(Krumm et al., 2014; Sanders et al., 2015), particularly for genes implicated via rare variants. However, several gene expression and epigenetic studies of ASD have implicated immune function in both brain tissue(Ander, Barger, Stamova, Sharp, & Schumann, 2015; Gupta et al., 2014; Nardone et al., 2014; Voineagu et al., 2011) and peripheral blood(Jalbrzikowski et al., 2015; Kong et al., 2013). Our results are concordant with these expression and epigenetic studies but still suggest a role for genetic variation in contributing to immune dysregulation in ASD, through SNP control of DNAm.

Beyond genome and epigenome enrichment analyses, specific meQTL targets also helped to "expand" ASD GWAS-implicated regions to include CpG sites, and their associated genes. While this does not increase or decrease statistical support for a particular GWAS SNP finding, better characterization of the functional architecture of the region can inform follow-up analyses of these hits. Two GWAS loci displayed evidence of meQTLs in peripheral blood, cord blood, and fetal brain, and many more loci displayed evidence of meQTLs in at least one tissue. These target CpG sites, and the genes they implicate, would not be identified via traditional genetic (i.e.

GWAS) analyses, since the sequence itself does not show ASD-related variability in these areas. Insights emerge only through the integration of SNP and DNAm data. Current PGC-AUT GWAS results are likely underpowered to provide reliable genome-wide hits. As larger GWAS of ASD emerge with higher-confidence findings, this cross-tissue meQTL mapping approach should be used to expand regions for follow-up, as recently demonstrated for schizophrenia in fetal brain(Hannon et al., 2016).

Finally, we sought to understand the propensity of meQTL targets, both generally and those controlled by psychiatric disorder-related SNPs, to overlap with regions of known functional activity. MeQTL targets of psychiatric SNPs in peripheral blood, fetal brain, and their intersection significantly overlapped with DHS sites, a result that is concordant with our observation of meQTL enrichment among ASD SNPs limited to peripheral blood and fetal brain. We also identified specific TFBSs enriched in psychiatric disorder meQTL targets such as TAF1 and STAT1. Recently, a study of nine families demonstrated both *de novo* and maternally inherited single nucleotide changes in *TAF1* to be associated with intellectual disability, facial dysmorphology, and neurological manifestations(O'Rawe et al., 2015). Our finding that binding sites for the TAF1 transcription factor overlap meQTL targets of psychiatric SNPs could serve a basis for future functional studies examining the link between *TAF1* mutations and adverse neurological phenotypes. Lastly, mutations in *STAT1* have been linked to early life combined immunodeficiency(Baris et al., 2016). The significant overlap with STAT1 TFBSs could thus serve as a starting point for functional work looking to understand the role of immune disorders in ASD and psychiatric phenotypes generally.

We also considered regulatory feature overlap with meQTL targets in general, for comparison to results for psychiatric meQTLs. Given the very large number of meQTL targets when not restricting to those downstream of psychiatric GWAS SNPs, many cell type specific DHS sites and chromatin marks showed significant enrichment for meQTLs in general. Among within-tissue analyses, the effect sizes were small. However, enrichment seen in meQTL targets that overlapped peripheral blood and cord blood, and meQTL targets that overlapped peripheral blood and fetal brain was of larger effect size (though still moderate). This is consistent with the idea that genetically controlled CpGs are more likely to have common function across tissues (given they all share genomic sequence) compared to those less under direct genetic control. For example, one study has demonstrated that cross-tissue meQTLs (the SNPs, rather than meQTL CpG targets) are enriched for miRNA binding sites(Smith et al., 2014). We have not found literature examining cross-tissue meQTL targets themselves. This finding suggests an important area for future research, that could give greater context to our and future investigations of cross-tissue meQTL targets in a disease context specifically.

In summary, our work is the first genome-wide study of meQTLs in the context of ASD to date. The results point to the utility of both brain and blood tissues in studies of ASD that integrate epigenetic data to enhance current GWAS findings for ASD. We show the utility of examining the meQTL targets of ASD SNPs in providing novel insights into functional roles like immune system processes that would not be apparent via genotype-based analysis in isolation. Our work suggests that genetic and epigenetic data integration, from a variety of tissues, will continue to provide ASD-related functional insights as GWAS findings and meQTL mapping across a variety of tissues improve.

## 2.6 References


1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., … McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Abdolmaleky, H. M., Zhou, J.-R., & Thiagalingam, S. (2015). An update on the epigenetics of psychotic diseases and autism. *Epigenomics*, *7*(3), 427–449. https://doi.org/10.2217/epi.14.85

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. https://doi.org/10.1038/13810

Ander, B. P., Barger, N., Stamova, B., Sharp, F. R., & Schumann, C. M. (2015). Atypical miRNA expression in temporal cortex associated with dysregulation of immune, cell cycle, and other pathways in autism spectrum disorders. *Molecular Autism*, *6*, 37. https://doi.org/10.1186/s13229-015-0029-9

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics (Oxford, England)*, *30*(10), 1363–1369. https://doi.org/10.1093/bioinformatics/btu049

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism*, *8*, 21. https://doi.org/10.1186/s13229-017-0137-9

Bakulski, K. M., Halladay, A., Hu, V. W., Mill, J., & Fallin, M. D. (2016). Epigenetic Research in Neuropsychiatric Disorders: the "Tissue Issue." *Current Behavioral Neuroscience Reports*, *3*(3), 264–274. https://doi.org/10.1007/s40473-016-0083-4

Baris, S., Alroqi, F., Kiykim, A., Karakoc-Aydiner, E., Ogulur, I., Ozen, A., … Barlan, I. B. (2016). Severe Early-Onset Combined Immunodeficiency due to Heterozygous Gain-of-Function Mutations in STAT1. *Journal of Clinical Immunology*. https://doi.org/10.1007/s10875-016-0312-3

Berko, E. R., Suzuki, M., Beren, F., Lemetre, C., Alaimo, C. M., Calder, R. B., … Greally, J. M. (2014). Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genetics*, *10*(5), e1004402. https://doi.org/10.1371/journal.pgen.1004402

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., … Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Chen, Y., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., … Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, *8*(2), 203–209. https://doi.org/10.4161/epi.23470

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., … Bolton, P. (2015). Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry*, *72*(5), 415–423. https://doi.org/10.1001/jamapsychiatry.2014.3028

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet (London, England)*, *381*(9875), 1371–1379. https://doi.org/10.1016/S0140-6736(12)62129-1

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., … International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994. https://doi.org/10.1038/ng.2711

Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., … Mill, J. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biology*, *13*(6), R43. https://doi.org/10.1186/gb-2012-13-6-r43

Davis, L. K., Gamazon, E. R., Kistner-Griffin, E., Badner, J. A., Liu, C., Cook, E. H., … Cox, N. J. (2012). Loci nominally associated with autism from genome-wide analysis show enrichment of brain expression quantitative trait loci but not lymphoblastoid cell line expression quantitative trait loci. *Molecular Autism*, *3*(1), 3. https://doi.org/10.1186/2040-2392-3-3

Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods*, *10*(1), 5–6. https://doi.org/10.1038/nmeth.2307

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–210.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Estes, M. L., & McAllister, A. K. (2015). Immune mediators in the brain and peripheral tissues in autism spectrum disorder. *Nature Reviews. Neuroscience*, *16*(8), 469–486. https://doi.org/10.1038/nrn3978

Feinberg, J. I., Bakulski, K. M., Jaffe, A. E., Tryggvadottir, R., Brown, S. C., Goldman, L. R., … Feinberg, A. P. (2015). Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *International Journal of Epidemiology*, *44*(4), 1199–1210. https://doi.org/10.1093/ije/dyv028

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., … Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, *15*(12), 503. https://doi.org/10.1186/s13059-014-0503-2

Gamazon, E. R., Badner, J. A., Cheng, L., Zhang, C., Zhang, D., Cox, N. J., … Liu, C. (2013). Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Molecular Psychiatry*, *18*(3), 340–346. https://doi.org/10.1038/mp.2011.174

Gauderman, W. J. (2002). Sample Size Requirements for Association Studies of Gene-Gene Interaction. *American Journal of Epidemiology*, *155*(5), 478–484. https://doi.org/10.1093/aje/155.5.478

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., … Arking, D. E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature Communications*, *5*, 5748. https://doi.org/10.1038/ncomms6748

Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*, *10*(11), 1024–1032. https://doi.org/10.1080/15592294.2015.1100786

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., … Mill, J. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nature Neuroscience*, *19*(1), 48–54. https://doi.org/10.1038/nn.4182

Horsthemke, B., & Wagstaff, J. (2008). Mechanisms of imprinting of the Prader-Willi/Angelman region. *American Journal of Medical Genetics. Part A*, *146A*(16), 2041–2052. https://doi.org/10.1002/ajmg.a.32364

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*(8), 955–959. https://doi.org/10.1038/ng.2354

IlluminaHumanMethylation450kmanifest. (n.d.). Retrieved August 5, 2016, from http://bioconductor.org/packages/IlluminaHumanMethylation450kmanifest/

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216–221. https://doi.org/10.1038/nature13908

Jalbrzikowski, M., Lazaro, M. T., Gao, F., Huang, A., Chow, C., Geschwind, D. H., … Bearden, C. E. (2015). Transcriptome Profiling of Peripheral Blood in 22q11.2 Deletion Syndrome Reveals Functional Pathways Related to Psychosis and Autism Spectrum Disorder. *PloS One*, *10*(7), e0132542. https://doi.org/10.1371/journal.pone.0132542

James, S. J., Shpyleva, S., Melnyk, S., Pavliv, O., & Pogribny, I. P. (2014). Elevated 5-hydroxymethylcytosine in the Engrailed-2 (EN-2) promoter is associated with increased gene expression and decreased MeCP2 binding in autism cerebellum. *Translational Psychiatry*, *4*, e460. https://doi.org/10.1038/tp.2014.87

Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., & de Bakker, P. I. W. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics (Oxford, England)*, *24*(24), 2938–2939. https://doi.org/10.1093/bioinformatics/btn564

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, *32*(Database issue), D493-496. https://doi.org/10.1093/nar/gkh103

Kong, S. W., Shimizu-Motohashi, Y., Campbell, M. G., Lee, I. H., Collins, C. D., Brewster, S. J., … Kunkel, L. M. (2013). Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. *Neurogenetics*, *14*(2), 143–152. https://doi.org/10.1007/s10048-013-0363-z

Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences*, *37*(2), 95–105. https://doi.org/10.1016/j.tins.2013.11.005

Ladd-Acosta, C., Hansen, K. D., Briem, E., Fallin, M. D., Kaufmann, W. E., & Feinberg, A. P. (2014). Common DNA methylation alterations in multiple brain regions in autism. *Molecular Psychiatry*, *19*(8), 862–871. https://doi.org/10.1038/mp.2013.114

Loke, Y. J., Hannan, A. J., & Craig, J. M. (2015). The Role of Epigenetic Change in Autism Spectrum Disorders. *Frontiers in Neurology*, *6*, 107. https://doi.org/10.3389/fneur.2015.00107

McClay, J. L., Shabalin, A. A., Dozmorov, M. G., Adkins, D. E., Kumar, G., Nerella, S., … van den Oord, E. J. C. G. (2015). High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biology*, *16*, 291. https://doi.org/10.1186/s13059-015-0842-7

missMethyl. (n.d.). Retrieved August 5, 2016, from http://bioconductor.org/packages/missMethyl/

Montano, C., Taub, M. A., Jaffe, A., Briem, E., Feinberg, J. I., Trygvadottir, R., … Feinberg, A. P. (2016). Association of DNA Methylation Differences With Schizophrenia in an Epigenome-Wide Association Study. *JAMA Psychiatry*, *73*(5), 506–514. https://doi.org/10.1001/jamapsychiatry.2016.0144

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational Psychiatry*, *4*, e433. https://doi.org/10.1038/tp.2014.70

Newschaffer, C. J., Croen, L. A., Fallin, M. D., Hertz-Picciotto, I., Nguyen, D. V., Lee, N. L., … Shedd-Wise, K. M. (2012). Infant siblings and the investigation of autism risk factors. *Journal of Neurodevelopmental Disorders*, *4*(1), 7. https://doi.org/10.1186/1866-1955-4-7

Nguyen, A., Rauch, T. A., Pfeifer, G. P., & Hu, V. W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *24*(8), 3036–3051. https://doi.org/10.1096/fj.10-154484

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., … Mandel, J. L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science (New York, N.Y.)*, *252*(5009), 1097–1102.

O'Rawe, J. A., Wu, Y., Dörfel, M. J., Rope, A. F., Au, P. Y. B., Parboosingh, J. S., … Lyon, G. J. (2015). TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations. *American Journal of Human Genetics*, *97*(6), 922–932. https://doi.org/10.1016/j.ajhg.2015.11.005

O'Roak, B. J., Stessman, H. A., Boyle, E. A., Witherspoon, K. T., Martin, B., Lee, C., … Eichler, E. E. (2014). Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nature Communications*, *5*, 5595. https://doi.org/10.1038/ncomms6595

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., … Scherer, S. W. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, *94*(5), 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

PriorityPruner. (n.d.). Retrieved August 5, 2016, from http://prioritypruner.sourceforge.net/

Psychiatric Genomics Consortium. (n.d.). Retrieved August 5, 2016, from http://www.med.unc.edu/pgc

Rapin, I. (1997). Autism. *The New England Journal of Medicine*, *337*(2), 97–104. https://doi.org/10.1056/NEJM199707103370206

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330. https://doi.org/10.1038/nature14248

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., … State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA*, *311*(17), 1770–1777. https://doi.org/10.1001/jama.2014.4144

Schendel, D. E., Diguiseppi, C., Croen, L. A., Fallin, M. D., Reed, P. L., Schieve, L. A., … Yeargin-Allsopp, M. (2012). The Study to Explore Early Development (SEED): a multisite epidemiologic study of autism by the Centers for Autism and Developmental Disabilities Research and Epidemiology (CADDRE) network. *Journal of Autism and Developmental Disorders*, *42*(10), 2121–2140. https://doi.org/10.1007/s10803-012-1461-8

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics (Oxford, England)*, *28*(10), 1353–1358. https://doi.org/10.1093/bioinformatics/bts163

Shi, J., Marconett, C. N., Duan, J., Hyland, P. L., Li, P., Wang, Z., … Landi, M. T. (2014). Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nature Communications*, *5*, 3365. https://doi.org/10.1038/ncomms4365

Smith, A. K., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., … Conneely, K. N. (2014). Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*, *15*, 145. https://doi.org/10.1186/1471-2164-15-145

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

sva. (n.d.). Retrieved August 5, 2016, from http://bioconductor.org/packages/sva/

Touleimat, N., & Tost, J. (2012). Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*, *4*(3), 325–341. https://doi.org/10.2217/epi.12.21

Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W., & Siegmund, K. D. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research*, *41*(7), e90. https://doi.org/10.1093/nar/gkt090

van Eijk, K. R., de Jong, S., Strengman, E., Buizer-Voskamp, J. E., Kahn, R. S., Boks, M. P., … Ophoff, R. A. (2015). Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood. *European Journal of Human Genetics: EJHG*, *23*(8), 1106–1110. https://doi.org/10.1038/ejhg.2014.245

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., … Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–384. https://doi.org/10.1038/nature10110

Wong, C. C. Y., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., … Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry*, *19*(4), 495–503. https://doi.org/10.1038/mp.2013.41

**Table 2.1: Descriptive characteristics, meQTL query parameters, and meQTL summary results for 4 tissues examined**

| | Sample Size | Meth SD Cutoff [b] | SNP MAF Threshold [c] | Max SNP to CpG Distance[d] | meQTL P-value thresholds [e] | # of meQTLs identified | # of meQTL targets identified |
|---|---|---|---|---|---|---|---|
| **Fetal Brain[a]** | 166 | NA | 5% | 1 Mb | 1.0e-08[f] | 299,992[f] | 7,863[f] |
| **Peripheral Blood** | 339 | 0.15 | 2.75% | 1 Mb | 3.1e-05[g]<br>1.0e-05[h]<br>3.0e-07[i] | 2,003,443[g]<br>1,878,577[h]<br>1,598,033[i] | 95,195[g]<br>85,250[h]<br>68,860[i] |
| **Cord Blood** | 121 | 0.15 | 7% | 500 Kb | 8.5e-06[g]<br>2.7e-06[h]<br>2.0e-07[i] | 1,374,554[g]<br>1,252,498[h]<br>1,032,370[i] | 41,681[g]<br>35,905[h]<br>28,423[i] |
| **Lung[a]** | 210 | NA | 3% | 500 Kb | 4.0e-05[h] | 22,866[h] | 34,304[h] |

[a]Publicly available data. [b]The probe standard deviation across samples that was used as an inclusion criterion for CpG sites in the meQTL query (blood datasets only). [c]The MAF cutoff used as an inclusion criterion for SNPs in the meQTL query. [d]The maximum distance between the SNP and CpG site used in the meQTL query for the peripheral blood, cord blood, and lung datasets, and the value at which results for filtered in the fetal brain results. [e]SNP to CpG association *P*-values considered in subsequent analyses. [f]FDR not specified. [g]FDR = 10% [h]FDR = 5% [i]FDR = 1%

**Table 2.2: Enrichment statistics for meQTLs derived from 4 tissue types in ASD GWAS SNPs**

| | **ASD *P*-value = 1e-03** | | | **ASD *P*-value = 1e-04** | | |
|---|---|---|---|---|---|---|
| | ***meQTL P-value = 1e-08*** | | | ***meQTL P-value = 1e-08*** | | |
| **Fetal Brain[1]** | 1.70 (<0.001) | | | 3.55 (<0.001) | | |
| | ***meQTL FDR = 10%*** | ***meQTL FDR = 5%*** | ***meQTL FDR = 1%*** | ***meQTL FDR = 10%*** | ***meQTL FDR = 5%*** | ***meQTL FDR = 1%*** |
| **Peripheral Blood[2]** | 1.22 (< 0.001) | 1.20 (< 0.001) | 1.23 (< 0.001) | 1.31 (0.001) | 1.40 (< 0.001) | 1.58 (< 0.001) |
| **Cord Blood[2]** | 1.14 (0.032) | 1.21 (0.011) | 1.20 (0.023) | 1.13 (0.299) | 1.10 (0.392) | 1.10 (0.406) |
| **Lung[1]** | - | 1.09 (0.343) | - | - | 0.80 (0.301) | - |

Enrichment fold statistics and *P*-values based on 1,000 permutations. [1]LD pruning performed with 1000 Genomes CEU samples. [2]LD pruning performed with study-specific genotype data. See Methods for additional details.

**Table 2.3: Gene Ontology terms significantly enriched in multiple tissue types in comparison of ASD-related meQTL targets to meQTL targets generally**

| Term | Peripheral Blood Scaled Rank[1] | Cord Blood Scaled Rank[1] | Fetal Brain Scaled Rank[1] |
|---|---|---|---|
| *response to interferon-gamma* | 0.14 | 0.11 | 0.11 |
| *positive regulation of relaxation of cardiac muscle* | 0.20 | 0.46 | 0.30 |
| *production of molecular mediator of immune response* | 0.65 | 0.22 | 0.28 |
| *cellular response to interferon-gamma* | NA | 0.07 | 0.09 |
| *detection of bacterium* | NA | 0.18 | 0.06 |
| *detection of biotic stimulus* | NA | 0.26 | 0.04 |
| *T-helper 1 type immune response* | NA | 0.08 | 0.34 |
| *regulation of interleukin-10 secretion* | NA | 0.09 | 0.43 |
| *interferon-gamma production* | NA | 0.57 | 0.19 |
| *regulation of interleukin-4 production* | NA | 0.24 | 0.62 |
| *interleukin-4 production* | NA | 0.29 | 0.60 |
| *interleukin-10 production* | NA | 0.25 | 0.74 |
| *tongue development* | NA | 0.68 | 0.32 |
| *inflammatory response to antigenic stimulus* | NA | 0.32 | 0.81 |
| *endochondral bone growth* | NA | 0.71 | 0.53 |
| *antigen processing and presentation of peptide or polysaccharide antigen via MHC class II* | 0.01 | 0.05 | NA |
| *T cell costimulation* | 0.05 | 0.01 | NA |
| *positive regulation of hormone secretion* | 0.09 | 0.04 | NA |
| *antigen receptor-mediated signaling pathway* | 0.08 | 0.13 | NA |
| *immunoglobulin production involved in immunoglobulin mediated immune response* | 0.24 | 0.03 | NA |
| *single organismal cell-cell adhesion* | 0.23 | 0.12 | NA |
| *single organism cell adhesion* | 0.34 | 0.16 | NA |
| *negative regulation of nonmotile primary cilium assembly* | 0.16 | 0.39 | NA |
| *antigen processing and presentation of polysaccharide antigen via MHC class II* | 0.02 | 0.58 | NA |
| *polysaccharide assembly with MHC class II protein complex* | 0.03 | 0.59 | NA |
| *protein-carbohydrate complex subunit organization* | 0.04 | 0.61 | NA |
| *microtubule sliding* | 0.29 | 0.38 | NA |
| *MHC protein complex assembly* | 0.06 | 0.75 | NA |
| *negative regulation of serine-type peptidase activity* | 0.41 | 0.41 | NA |
| *regulation of satellite cell activation involved in skeletal muscle regeneration* | 0.39 | 0.45 | NA |
| *protein repair* | 0.43 | 0.43 | NA |
| *regulation of serine-type peptidase activity* | 0.48 | 0.47 | NA |

| | | | |
|---|---|---|---|
| *protein localization to basolateral plasma membrane* | 0.46 | 0.55 | NA |
| *lymphocyte migration into lymphoid organs* | 0.47 | 0.62 | NA |
| *Peyer's patch morphogenesis* | 0.60 | 0.70 | NA |
| *regulation of homeostatic process* | 0.45 | 0.92 | NA |
| *skeletal muscle satellite cell activation* | 0.88 | 0.63 | NA |

[1] Scaled rank refers to enrichment *P*-value based rank divided by the number of marginally significant terms post REVIGO filtering for that tissue (peripheral blood: 95, cord blood: 76, fetal brain: 47). 'NA' shown for terms that did not appear in these lists for that tissue. Terms are lumped into sections based on cross-tissue overlap: Section 1 – all three tissues, Section 2 – cord blood and fetal brain, Section 3 – peripheral blood and cord blood. Within each of these sections, terms are arranged according to the sum of the scaled ranks. See Methods for additional details.

**Table 2.4: Samples downloaded from Roadmap Epigenomics Project for 5 different histone modifications.**

| | H3K27me3 | H3K36me3 | H3K4me1 | H3K4me3 | H3K9me3 |
|---|---|---|---|---|---|
| Adult Lung | NA | GSM1059437 | GSM1059443 | GSM1227065 | GSM1120355 |
| | GSM1220283 | GSM956014 | GSM910572 | GSM915336 | GSM906411 |
| | | | | | |
| Fetal Brain | GSM621393 | GSM621410 | GSM706850 | GSM621457 | GSM621427 |
| | GSM916061 | | | | GSM916054 |
| | | | | | |
| Peripheral Blood | GSM1127130 | GSM1127131 | GSM1127143 | GSM1127126 | GSM1127133 |
| | GSM1127142 | GSM613880 | | | GSM613878 |

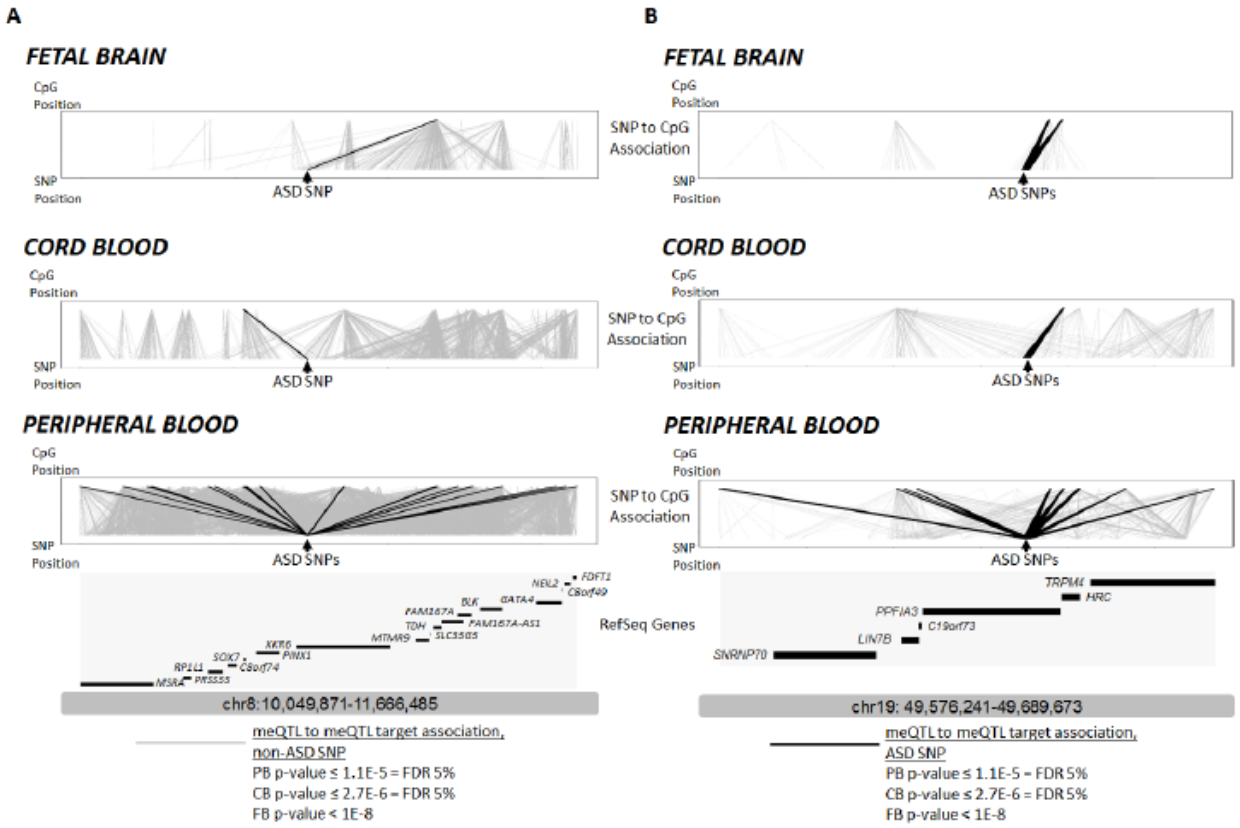



**Figure 2.1: 'Expansion' of ASD loci through meQTL mapping in peripheral blood, cord blood, and fetal brain** Each tissue-specific panel presents, from bottom to top: genomic location, gene annotations, SNP locations, SNP-CpG associations, CpG locations. Light gray meQTL association lines denote all SNP to CpG associations in that tissue type; Dark meQTL association lines denote SNP-CpG associations for ASD-associated SNPs in PGC (*P*-value <= 1e-04). Panel A) Locus at chr8; Panel B) Locus at chr19. Data are presented for meQTL maps for fetal brain (top); cord blood meQTLs (middle), and peripheral blood meQTLs (bottom). Please note locus coordinates differ from those in **Supplementary Data II-6** because in this context they encompass the locations of meQTL target CpG sites.

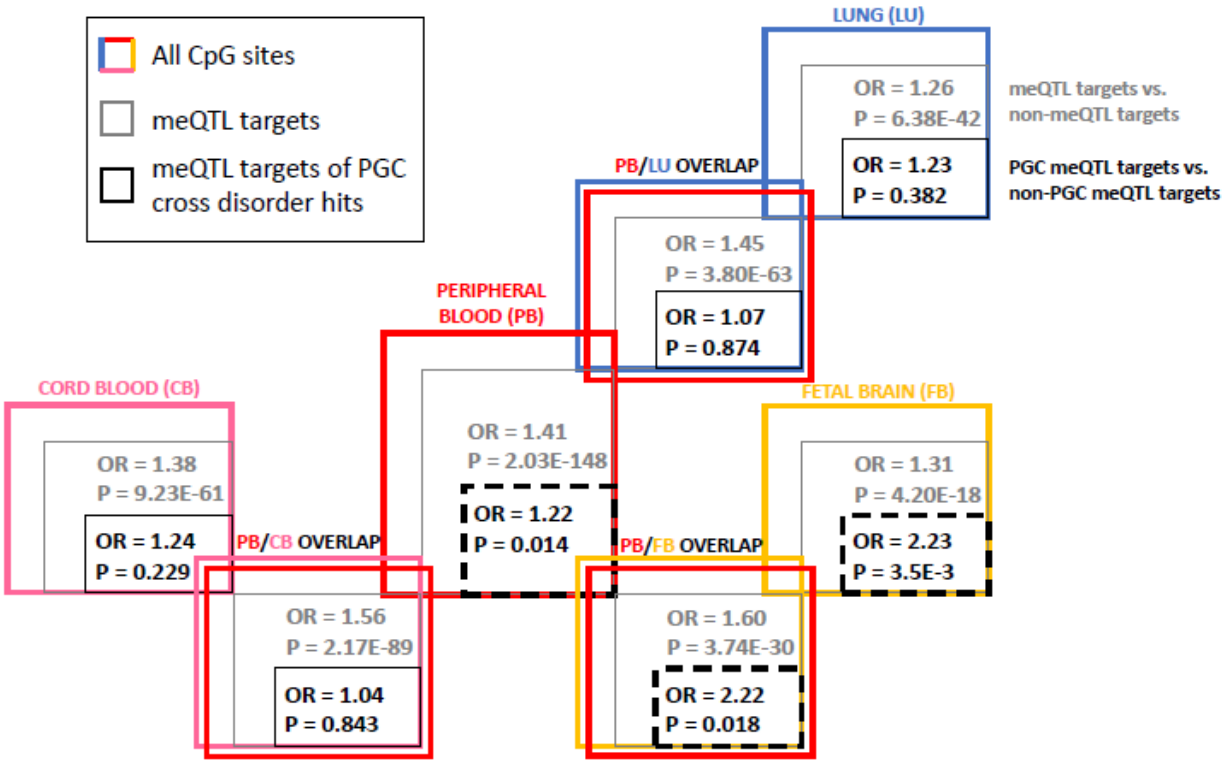



**Figure 2.2: Enrichment of meQTL target CpG sites in DNaseI hypersensitive sites** We identified the meQTL targets (at FDR 5% in peripheral blood, cord blood, and lung and past 1e-08 *P*-value threshold in fetal brain results) in peripheral blood, cord blood, fetal brain and lung as well those meQTL targets that were present in the overlap of peripheral blood with the other three tissues. Odds ratio and *P*-value in gray text represent enrichment fold statistic and *P*-value from Fisher's exact tests examining the tendency of meQTL targets to overlap with DHS sites compared to CpG sites that were not meQTL targets. Odds ratio and *P*-value in black text represent enrichment fold statistic and *P*-value from Fisher's exact tests examining the tendency of meQTL targets of significant (*P*-value <= 1e-04) SNPs from the PGC cross-disorder results or their proxies ($r^2$ >= 0.8) to overlap with DHS sites compared to CpG sites that were not meQTL targets of the same SNPs. A full list of enrichment statistics and *P*-value for both tests against a total of 181 cell-type specific DHS sites, cell-type specific chromatin marks, and transcription factor binding sites is available in **Supplementary Data II-7 and II-8**.

# CHAPTER 3: FUNCTIONAL ANNOTATION OF GWAS FINDINGS, ACROSS AFFECTED AND SURROGATE TISSUE TYPES, TO PROVIDE MECHANISTIC INSIGHT IN AUTISM SPECTRUM DISORDER

## 3.1 Abstract

Background: Common variant GWAS discovery studies for autism spectrum disorder (ASD) have begun to realize their potential with recent substantial increases in sample size through the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) initiative and Psychiatric Genomics Consortium ASD workgroup (PGC-AUT). Genome-wide significant common variants have been identified as well as a number of loci with suggestive significance. However, the downstream functional biology impacted by ASD associated common variants remains elusive. A separate but growing literature supports epigenetic involvement in the etiology of ASD. Thus, studies that integrate genome-wide association study (GWAS) findings for ASD and epigenetics provide a fruitful avenue for furthering our understanding of altered biologic mechanisms in ASD. In fact, we recently demonstrated the utility of meQTL maps for uncovering ASD biology that otherwise would not have been implicated by GWAS findings alone. Here, we extend this framework and assess meQTLs in a new tissue relevant to ASD – placenta – to uncover downstream biology related to the most recent PGC-AUT GWAS results.

Objectives: This study aims to expand upon our previous study that integrated genetic and epigenetic data to inform our understanding of biological mechanisms in ASD by using GWAS summary statistics from the largest ASD GWAS to date, nearly tripling the number of samples as the previous largest GWAS, and by generating new meQTL maps for placenta tissue. We then use the new GWAS results and placenta meQTL maps to: A) determine whether ASD associated SNPs are enriched for meQTLs and B) identify methylation loci and biologic pathways that are

controlled by ASD SNPs. We perform these analyses in tissues thought to be affected (e.g. fetal brain) as well as surrogate tissues (e.g. blood) in ASD.

Results: ASD-related SNPs ($P_{GWAS}$ < 1E-5) were enriched for placenta (fold enrichment = 1.25, $P_{Enrichment}$ = 0.011) meQTLs, as well as fetal brain (FE = 4.53, $P_{Enrichment}$ < 0.001), peripheral blood (FE = 1.39, $P_{Enrichment}$ = 0.011), and cord blood (FE = 2.16, $P_{Enrichment}$ < 0.001) meQTLs but not adipose tissue meQTLs (FE = 1.41, $P_{Enrichment}$ = 0.317). Gene ontology analysis revealed both tissue-specific pathways targeted by ASD SNPs, such as glial cell development in fetal brain and virus-host interaction in placenta, as well as endopeptidase activity and mitochondrion transport pathways that were commonly engaged by ASD SNPs across fetal brain, placenta, cord blood, and peripheral blood tissue types. Finally, all loci containing the five genome-wide significant ASD GWAS SNPSs contain meQTLs in placenta, including two that are also contain peripheral blood meQTLs. Mapping the locations of their downstream CpG site targets identified additional putative ASD candidate genes in specific tissue types.

Conclusion: Integration of ASD GWAS summary statistics and tissue-specific meQTL maps provide insights into ASD biology as well as into the utility of surrogate tissues for ASD etiology based research. SNP-based enrichment analyses show the broad function of ASD SNPs through DNAm across brain and peripheral tissue types, and CpG-based enrichment analyses illuminate the biological pathways through which these SNPs function, demonstrating their downstream impact. Thus our study provides needed characterization of functional biology impacted by ASD SNPs and adds further evidence for the epigenetic basis to ASD. Future studies of epigenetics in ASD across tissue types should continue to integrate genotype data for etiologic insights.

### 3.2 Introduction

Autism spectrum disorder (ASD) is characterized by social, behavioral, and communication deficits and, as of 2012, affects 1 in 68 children in the US (Christensen et al., 2016). ASD is known to have a strong genetic basis (Colvert et al., 2015; Gaugler et al., 2014; Sandin et al., 2014). Most genetic studies of ASD to date have aimed to identify rare variants, including inherited and *de novo* single nucleotide and copy number variants (CNVs), associated with the disorder (Iossifov et al., 2014; O'Roak et al., 2014; Sanders et al., 2015). Common variant findings are increasingly being realized through the Psychiatric Genomics Consortium ASD working group (PGC-AUT). The most recent meta-analysis results nearly tripled the number of samples analyzed and resulted in the identification of 5 genome-wide significant SNPs (Grove et al., 2017). However, the functional biology implicated by common variation in ASD is not well understood. Previous studies of neuropsychiatric phenotypes have sought to elucidate these mechanisms by integrating disorder-specific GWAS results along with data for expression quantitative trait loci, or eQTLs (Davis et al., 2012; Gamazon et al., 2013; van Eijk et al., 2015). Epigenetic factors, such as DNA methylation (DNAm), can be integrated with GWAS data in a similar manner, given that SNPs can control DNAm levels locally and distally as with expression loci (Bell et al., 2011). These studies have been undertaken in bipolar disorder (Gamazon et al., 2013) and schizophrenia (Hannon et al., 2016; van Eijk et al., 2015), specifically demonstrating an enrichment of methylation quantitative trait loci (meQTLs), or SNPs that control DNAm, in disorder-associated SNPs.

The evidence for a role of epigenetic factors such as DNAm in the etiology of ASD is increasingly developing. ASD epigenetic studies are founded upon the fact that epigenetic mechanisms have been implicated by ASD rare variant findings (Krumm, O'Roak, Shendure, &

Eichler, 2014; Pinto et al., 2014; Sanders et al., 2015), and that many disorders (Rett, Angelman, Fragile X) that overlap with ASD phenotypically have an epigenetic basis (Budimirovic & Kaufmann, 2011; Jedele, 2007; Mount, Charman, Hastings, Reilly, & Cass, 2003). Studies in postmortem brain samples have shown ASD-based changes in DNAm (Gregory et al., 2009; James, Shpyleva, Melnyk, Pavliv, & Pogribny, 2013; Ladd-Acosta et al., 2014; Nagarajan, Hogart, Gwye, Martin, & LaSalle, 2006; Nardone et al., 2014; Zhu et al., 2014) and histone acetylation (Sun et al., 2016). Brain-based samples have limited accessibility. Therefore, many peripheral tissues have been examined in this context as well including peripheral blood (Aldinger, Plummer, & Levitt, 2013; Gregory et al., 2009; Wong et al., 2014), buccal epithelium (Berko et al., 2014), lymphoblastoid cell lines (Nguyen, Rauch, Pfeifer, & Hu, 2010), the sperm of fathers of children with ASD (Feinberg et al., 2015), and placenta (Schroeder et al., 2016). Therefore, there is evidence to suggest epigenetics may be involved in ASD etiology and that changes related to ASD may be detected in both affected and accessible tissue sources.

We recently integrated genetic and epigenetic data to better characterize the biology of disease associated common variants and DNAm in ASD (see Chapter 2 and (Andrews et al., 2017)). This work was based on results from the initial, considerably smaller, PGC-AUT GWAS that did not reach genome-wide significance (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017). A unique aspect of our approach was to leverage DNAm information from both affected and peripheral surrogate tissue types which enabled us to understand whether ASD meQTL biology was specific to brain tissue or whether some could also be detected in surrogate tissues. Here, we applied the same approach using the most recent PGC-AUT GWAS results (Grove et al., 2017), that now reach genome-wide significance, to

determine: (1) whether ASD-associated SNPs are enriched for meQTLs, (2) what methylation loci are controlled by the 5 genome-wide significant SNP and, thus, provide novel candidate genes and loci for follow-up functional work, and (3) identify pathways enriched in the methylation targets of ASD meQTLs. In addition, we expand upon our previous work in blood and brain tissues (Andrews et al., 2017) by generating and using new meQTL maps in a fourth tissue - placenta – that is relevant to ASD (see details in Chapter 1).

## 3.3 Methods

### *3.3.1 EARLI Study Description*

The Early Autism Risk Longitudinal Investigation (EARLI) is a prospective, enriched-risk birth cohort that has been described in detail elsewhere (Newschaffer et al., 2012). The EARLI study was approved by Human Subjects Institutional Review Boards (IRBs) from each of the four study sites (Johns Hopkins University, Drexel University, University of California Davis, and Kaiser Permanente). Informed consent was obtained from all participating families. The 232 mothers with a subsequent child born through this study had births between November 2009 and March 2012. Placental biopsy samples were collected after delivery at each clinical lab site using Baby Tischler Punch Biopsy Forceps. Samples were stored at ambient temperature in RNAlater vials (Qiagen, Cat. No. 76154) and shipped same-day to the Johns Hopkins Biological Repository (JHBR) in Baltimore, Maryland, for storage at -190°C until processing. After biopsy, the placentas were then placed flat in a bag with 10% formalin at a volume roughly equivalent to the size of the placenta.

### *3.3.2 Whole genome bisulfite sequencing, alignment, and quality control*

Genomic DNA (gDNA) from biopsy punches was extracted at JHBR using a QIAgen QIAsymphony automated workstation with the Blood 1000 protocol of the DSP DNA Midi kit (Cat. No. 937255) as per manufacturer's instructions. 250ng gDNA was bisulfite treated using the Zymo EZ-96 DNA Methylation Kit (Cat. No. D5004), and sequencing libraries were made with the NEBNext Ultra DNA Library Prep Kit for Illumina. Library preparation was performed at the Center for Epigenetics and libraries were sequenced at ~4x coverage at the High Throughput Sequencing Center at the Johns Hopkins University School of Medicine using the 3 Illumina HiSeq 2500 instruments. Initially 133 placenta samples were run on 16 flow cells; two samples failed due to a fluidics error and were re-run using a $17^{th}$ flow cell.

Raw .fastq files were aligned the human reference genome (hg38) downloaded from Ensembl (Aken et al., 2017) via Bismark (v0.16.3) (Krueger & Andrews, 2011) using the default arguments for paired-end sequencing data. M-bias plots (Hansen, Langmead, & Irizarry, 2012) were used to inform the trimming of the 5 bp of both the first and second reads, and methylation status was called using the Bismark methylation extractor function. All subsequent analyses used R (version 3.4.2) and the R package *bsseq* (version 1.13.6) ("bsseq," n.d.). The bsseq function *read.bismark()* was used to read in Bismark output files ("cytosine report" format) and obtain DNAm measurements on a per-CpG site basis. A total of 29,091,077 CpG sites were measured with at least 1 read in 1 sample out of the total 133. Next methylation levels were smoothed using the BSmooth algorithm that has been previously described (Hansen et al., 2012). Briefly, this technique employs a local-likelihood smoothing algorithm to recapitulate higher coverage whole genome bisulfite sequencing (WGBS) data from lower coverage data, exploiting known

local correlation structure in DNAm data. We used the *BSmooth()* function in the *bsseq* R package with the default arguments.

*3.3.3 Genotype data and methylation quantitative trait loci query*

Measurement, quality control, and imputation for the EARLI placenta genotype samples were described previously for their matched cord blood samples (Andrews et al., 2017). Overall, 10,773,289 SNPs were available on 121 out of 133 samples on which we had WGBS DNAm data. After imposing a 5% MAF threshold, 7,187,552 SNPs remained. Finally, we used the *liftOver* (Hinrichs et al., 2006) tool to convert the hg19 SNP positions into hg38 positions, to match the genome build to which the WGBS data were aligned. 5,155 SNPs failed this conversion process, leaving a final total of 7,182,397 SNPs on the 121 samples.

We took two steps to prepare the DNAm data for the meQTL query. First, in the total dataset of 29,091,077 CpG sites and 133 samples, we filtered out sites with standard deviations less than the $15^{th}$ percentile of the total vector of standard deviations calculated at every sample. This step was done to reduce computational burden by removing sites of low variability at which we are underpowered to discover meQTLs and/or meQTL targets. This step limited the number of available CpG sites to 24,727,415 sites. Next we imposed similar coverage filtering as done for the previous analyses. Specifically we limited the remaining data to retain sites with $\geq 2$ reads in more than two-thirds of the total number of samples, leaving 22,710,938 DNAm sites.

Pairwise associations between all SNPs and CpG sites within a distance of 1 Mb from each other were estimated via the *MatrixEQTL* R package (Shabalin, 2012). Percent methylation was regressed onto genotype using an additive model, adjusting for the first two principal components of ancestry and gestational age. FDR was controlled via the iterative hypothesis

weighting, implemented via the *IHW* R package. This method assigns p-value weights according to an informative covariate and increases power over typical FDR control methods (Ignatiadis, Klaus, Zaugg, & Huber, 2016). We used distance between SNP and CpG site as our specified covariate, as suggestive by the IHW authors and given previous evidence for its strong role in dictating meQTL associations (Andrews et al., 2017; Hannon et al., 2016; Shi et al., 2014).

### *3.3.4 Additional meQTL datasets*

We used the peripheral blood, cord blood and fetal brain meQTL (Hannon et al., 2016) lists from our previous ASD meQTL study (Andrews et al., 2017). We also utilized a list of cis-meQTLs from a previous adipose tissue study (Volkov et al., 2016) and obtained a list of SNPs that passed QC (N = 592,794) and CpG sites that passed QC (N = 477,891) from the study authors. These meQTL results were used in an analogous manner as the lung tissue meQTLs from our previous study (Andrews et al., 2017); they were included as a negative control given that adipose tissue is likely a non-ASD-related tissue.

### *3.3.5 Enrichment of meQTLs in ASD-associated SNPs*

We tested for enrichment of meQTL from five tissue types among ASD GWAS SNPs. ASD SNPs were obtained from the latest iteration of the PGC-AUT analysis based on 18,381 cases and 27,696 controls (Grove et al., 2017). This list consisted of a total of 9,112,386 SNPs, with 8,209 SNPs having an ASD association p-value < 1E-4, and 4,027 SNPs with p-value < 1E-5.

We implemented an identical SNP enrichment pipeline to our previous analysis (Andrews et al., 2017). This pipeline proceeds in the following steps: 1) Determination of SNP overlap between those used in the meQTL query and those used in the PGC analysis via direct or proxy overlap as determined via the SNAP software, 2) LD pruning of the remaining SNP set ‘supervised’ by

ASD p-value via the PriorityPruner software (v0.1.2) ("PriorityPruner," n.d.), 3) Running 1000 permutations employing matching of null SNP sets in equal number to the ASD SNP set on bins of MAF and the number of CpG sites in the meQTL discovery window. Full details on parameters used for each pipeline step can be found in our previous paper (Andrews et al., 2017). Finally, to explore the sensitivity of our results to definitions of a meQTL and an ASD-related SNP, we evaluated enrichment at two PGC-AUT p-value cutoffs (1E-4, 1E-5) and two meQTL p-value cutoffs (FDR = 5% and FDR = 1%) for peripheral blood, cord blood, and placenta. For the publically available adipose tissue and fetal brain meQTL lists were limited to the available information; therefore we only assessed our results at $P < 1E-8$ for fetal brain and $P < 4.8E-10$ for adipose tissue.

The LD pruning steps were conducted with sample-specific genotype data for in house data (peripheral blood, cord blood, placenta) and with 1000 Genomes CEU samples (1000 Genomes Project Consortium et al., 2012) when using the publically available fetal brain and adipose tissue meQTL lists.

*3.3.6 Gene ontology analysis of meQTL targets*

We next sought to assess which biological pathways are implicate via the location of meQTL targets of ASD-related SNPs. Therefore, we identified meQTL targets of ASD SNPs (and their proxies) and conducted a gene ontology analysis in which we compared these CpG sites against a background of meQTL targets as whole. The choice of this background set, rather than measured CpG sites, was so that we could enumerate the pathways specifically engaged by ASD SNP control of methylation and not SNP control more generally.

For all of the tissues in which DNAm data was measured via the 450K Array (peripheral blood, cord blood, fetal brain) we implemented an identical meQTL target enrichment pipeline as in our previous analysis (Andrews et al., 2017). First we filtered the list of meQTL targets to those associated with an ASD SNP (defined as GWAS p-value < 1E-5) or their proxies (defined as $r^2 >$ 0.8 within 500 kb window in CEU 1KG as defined via the SNAP software). Only meQTLs passing the respective meQTL FDR = 1% thresholds in peripheral blood and cord blood and $p <$ 1E-8 in fetal brain were used in this analysis. We again removed CpG sites overlapping SNPs within 10bp of the CpG site or at any single base extension, due to a previous analysis demonstrating the significant impact of these CpG sites in a functional-type enrichment analysis of CpG sites (McClay et al., 2015), and the fact that these sites were filtered out in the fetal brain study (Hannon et al., 2016). Finally we used the *gometh()* function in the *MissMethyl* R package, which maps CpG IDs to genes and accounts for non-uniform coverage of genes inherent to the 450K Array ("missMethyl," n.d.). We retained nominally significant ($p_{enrichment} < 0.05$) terms for cross-tissue overlap analyses.

For placenta, we employed the same meQTL target filtering as described above for the other tissues (using meQTLs defined at FDR = 1%). However, DNAm in placenta was measured via WGBS, which prevented the direct utilization of the same GO enrichment pipeline via *MissMethyl*, which was designed for Illumina methylation arrays. Therefore, we adapted the *gometh()* function for WGBS data by customizing the functionality to map CpG site positions to genes. Specifically, we downloaded a list of RefSeq genes (hg38) from the UCSC Genome Browser table (Karolchik et al., 2004), and defined the promoter regions of these genes as 2000 bp upstream of the listed gene transcription start site. We then used the *findOverlaps()* function

in the *GenomicRanges* R package (Lawrence et al., 2013) to find the genes (or associated promoters) that overlapped the placenta meQTL targets. After defining these gene lists for both placenta ASD meQTL targets and placenta meQTL targets generally, we used the enrichment test functionality of the *gometh()* function in *MissMethyl* to perform an analogous enrichment as done for the 450K Array-based studies that takes into account differential coverage of genes. Finally we again retained nominally significant terms ($p_{enrichment} < 0.05$) and identified different patterns of overlap with the respective terms from the other 3 tissues.

### *3.3.7 GWAS loci expansion via meQTL target information*

We sought to demonstrate the utility of meQTL target positional information to identify additional genes of interest in addition to those identified by ASD GWAS SNP positions alone. We used the same meQTL relationships as described in the GO meQTL target enrichment analysis. We again downloaded a RefSeq gene list from the UCSC Genome Browser (Karolchik et al., 2004); this list was in build hg19 to match the meQTL positional information of the peripheral blood, cord blood, and fetal brain data. We used the *liftOver()* tool (Hinrichs et al., 2006) to convert the hg38 positional information for the placenta DNAm data to hg19. We filtered this gene list to exclude long intergenic non-coding RNAs, long non-coding RNAs, microRNAs, and small associated RNAs. We also collapsed SNPs into bins by LD blocks defined via recombination hot spot data from 1000 Genomes (1000 Genomes Project Consortium et al., 2012, p. 0).

## 3.4 Results

### *3.4.1 Genome-wide meQTL mapping in placenta*

We conducted a genome-wide analysis to identify meQTLs and map CpG locus methylation under their control, i.e. their methylation targets, in placenta tissue using unified imputed genome-wide genotype data and WGBS-measured DNAm data from the same individuals. Using a liberal significance threshold (FDR = 5%), we identified a total of 5,022,957 meQTLs and 8,365,660 meQTL methylation targets. On average, each meQTL was associated with methylation levels at 96 CpG sites. Each CpG target was associated with 58 meQTL SNPs, on average. Using a more stringent significance threshold (FDR = 1%), we identified a total of 4,424,678 meQTLs and 5,466,372 meQTL methylation targets. On average, each meQTL was associated with 82 CpG sites and each CpG target was associated with 66 meQTL SNPs. We observed that the degree of significance, i.e. p-values, decays with increasing physical distance between a meQTL and its methylation target (**Supplementary Figure III-1**).

There were 2,620,313 overlapping SNPs considered for meQTL analysis in placenta, fetal brain, cord blood, and peripheral blood. Of these, 111,574 SNPs (4.26%) were identified as meQTLs in all 4 tissues, 902,884 SNPs (34.46%) were identified as meQTLs in placenta only, and 483,903 (18.47%) SNPs were identified as meQTLs in both placenta and peripheral blood (**Supplementary Table III-1**).

*3.4.2 ASD associated SNPs, identified via GWAS, are enriched for meQTLs*

Using the most recently released autism GWAS summary statistics from the PGC-AUT workgroup (Grove et al., 2017) and existing meQTL maps for fetal brain, peripheral blood, adipose, and cord blood tissues (see Chapter 2 and (Andrews et al., 2017)) as well as our newly generated placenta meQTL maps, we assessed whether ASD-associated SNPs are enriched for meQTLs among each of tissue type. Using the most stringent significance thresholds (a $P_{GWAS}$ <

1E-5 and meQTL p-value = 1.0E-08 for fetal brain or 4.8E-10 for adipose or meQTL FDR < 1% for cord blood, peripheral blood, and placenta), we observed significant meQTL enrichment in ASD SNPs in fetal brain, cord blood, peripheral blood, and placenta tissues (**Table 3.1**). Fetal brain displayed the strongest effect sizes (fold enrichment = 4.53), followed by in decreasing order cord blood (fold enrichment = 2.16), peripheral blood (fold enrichment = 1.39), and placenta (fold enrichment = 1.25). Similar results were observed using less stringent significance thresholds for these tissue types, although, the effect sizes decreased somewhat for fetal brain and placenta. In adipose tissue, significant enrichment was observed at the less stringent GWAS threshold (fold enrichment = 2.33, $P_{Enrichment} < 0.001$) but a smaller fold change and non-significant p-value was observed at the more stringent GWAS threshold (fold enrichment = 1.41, $P_{Enrichment} = 0.317$).

*3.4.3 Downstream biology implicated by pathway analysis of ASD meQTL methylation targets*

We determined whether the methylation target loci controlled by ASD meQTLs are enriched for specific biologic pathways. We compared ASD meQTL target pathways to a background set of target pathways implicated for meQTLs, more generally. We identified 137, 208, 143, and 157 gene ontology (GO) biological processes enriched, at nominal significance ($p < 0.05$), in ASD meQTL targets relative to meQTLs, generally, for fetal brain, peripheral blood, cord blood, and placenta, respectively (**Supplementary Tables III-2** to **III-5**).

A total of 6 biological processes overlapped across all 4 tissue types. These included one process involved in the apoptotic process, 2 processes involved in endopeptidase activity and 3 processes involved in mitochondrion transport. Several processes involving RNA 3'end processing and caveolin-mediated endocytosis overlapped 2 or 3 tissues, and the term "positive regulation of

neuron death" overlapped fetal brain and placenta (**Figure 3.2**). We also observed some interesting GO terms that were specific to one tissue type. For example, terms such as "glial cell development", "positive regulation of hormone secretion", and "memory" were specific to fetal brain (**Supplementary Table III-2**), many Wnt singling pathways emerged in the peripheral blood findings (**Supplementary Table III-3**), and "modulation by virus of host molecular function" and "syncytiotrophoblast cell differentiation involved in labyrinthine layer" were specific to the placenta (**Supplementary Table III-5**).

*3.4.4 Identification of novel downstream functional candidate genes for ASD*

We next sought to identify new candidate genes for ASD, that were not identified by the GWAS SNP results, using our meQTL maps in fetal brain, cord and peripheral blood, and placenta tissues. The most recent PGC-AUT GWAS (18,381 cases and 27,969 controls) revealed 4,027 ASD-associated (P < 1E-5) SNPs; 2,173 of those were a meQTL in all 4 - peripheral blood, cord blood, fetal brain and placenta - tissues (**Supplementary Table III-6**). We found that the methylation targets of these meQTL ASD-associated SNPs were often located in loci and genes outside of the LD block of the ASD-associated SNP itself; thus, revealing new candidate functional targets and genes related to ASD that were not implicated by the GWAS SNPs themselves. For example, one genome-wide significant ASD meQTL SNP is harbored in a locus where ASD-related SNPs (P < 1E-5) are located in *C8orf74* but control methylation levels at a different locus located in *PINX1* in both peripheral blood and placenta tissues (**Figure 3.2**). Of the 4 loci harboring additional genome-wide significant SNPs identified in the PGC-AUT GWAS analysis, one contained peripheral blood meQTLs and all 4 contained placenta meQTLs (**Table 3.2** and **Supplementary Figures III-2** to **III-5**); they were not detected as meQTLs in

the other tissues we tested. In 3 of these cases (rs910805, rs71190156, and rs201910565, see **Figures III-2**, **III-3** and **III-5**), the actual genome-wide significant SNP is a meQTL itself. For example, placenta meQTL maps revealed the most genome-wide significant SNP (rs910805), an intergenic SNP, controls methylation levels in the placenta at *KIZ*, *KIZ-AS1*, and *NKX2-4* (**Figure III-2**).

### 3.5 Discussion

In this work, we extended our previous framework and results integrating ASD GWAS results and meQTL maps to assess epigenetic targets of ASD SNPs in a new tissue type relevant to autism, placenta, using summary statistics from the largest ASD GWAS study to date with nearly triple the number of samples compared to the previous iteration (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017). We found that ASD-related SNPs were enriched for meQTLs from peripheral blood, cord blood, fetal brain, and placenta. Pathway analyses of the CpG methylation targets of ASD SNPs revealed enrichment of endopeptidase and mitochondrion pathways across fetal brain, blood and placenta tissues. However, most enriched ASD target pathways showed tissue specificity. Finally, our meQTL maps implicated new ASD candidate loci and genes for future follow-up studies that were located beyond those implicated by the SNP positions alone.

This expansion of our previous result, which included a new tissue type and the latest GWAS results from a substantially larger sample size, replicated our previous finding showing an enrichment of peripheral blood and fetal brain meQTLs in ASD GWAS hits (Andrews et al., 2017), and additionally demonstrated this same relationship with placenta meQTLs. We also observed this same enrichment in cord blood meQTLs, a finding we did not observe in our

previous study that utilized the previous iteration of the PGC-AUT GWAS results for analysis. These results broadly implicate fetal brain, cord blood, peripheral blood from early life, and the placenta as key tissues through which genetic control of DNAm is relevant in an ASD context. These findings warrant the continued investigation of DNAm signatures of ASD in these tissue types (i.e. differentially methylated positions or regions related to ASD), particularly at DNAm locations known to be controlled by ASD-related SNPs.

Our placenta studies herein were particularly unique in that we derived placenta meQTLs using WGBS-measured DNAm, which offers drastically improved coverage of the DNA methylome over the commonly used 450K array. This fact should be taken into account when interpreting our results. For example, the fact that every genome-wide significant locus from the PGC-AUT contained a placenta meQTL is more of a reflection of the greater opportunity for SNPs to be called meQTLs in placenta (owing to greater number of measured CpGs in the meQTL detection window) than it is a greater engagement of the placenta methylome downstream of ASD SNPs. Similarly, our cross-tissue meQTL results revealed a great number of meQTLs that were identified in placenta but not seen in fetal brain, cord blood, or peripheral blood. This fact – that WGBS-measured DNAm leads to inherently greater meQTL findings in placenta – is necessarily true even given our use of strict multiple testing thresholds in the meQTL query (Ignatiadis et al., 2016), and points to larger issues that complicate cross-tissue meQTL comparisons such as differing SNP and/or CpG variability thresholds used as inclusion criteria, meQTL detection window size, and DNAm measurement platform, as we have previously discussed (Andrews et al., 2017). Therefore, cross-tissue comparisons with these placenta data will be particularly of interest as WGBS-measured DNAm becomes available for fetal brain and blood tissues.

The use of adipose tissue meQTLs in this current study functioned as a negative control for our SNP-based enrichment analyses. We observed significant enrichment in adipose tissue meQTLs at the more liberal GWAS threshold examined to define an ASD-related SNP but not at the more stringent threshold of an ASD-related SNP. This pattern was not consistent with the other tissues we examined. Observed adipose tissue meQTL enrichment at the more liberal GWAS threshold may indicate a possible role for adipose tissue in ASD biology, or, more likely, a less tissue-specific enrichment for meQTLs when including SNPs more loosely-associated with ASD. Therefore, characterization of meQTL relationships in an ASD context are likely more meaningful at the GWAS threshold where we did not observe adipose tissue meQTL enrichment (i.e. GWAS p-value 1E-5). In light of this, GO enrichment analyses and GWAS loci expansion analyses took place using meQTL relationships defined via this more stringent GWAS threshold.

In our previous study, we reported that the CpG site targets of ASD-related SNPs were engaged in immune response pathways across peripheral blood, cord blood, and fetal brain (Andrews et al., 2017). This finding was particularly intriguing given the known comorbidity of ASD and immune-related phenotypes, and the fact that many gene expression and DNAm-based studies of ASD have identified immune response pathways. In this study, we performed similar pathway analyses but did not replicate the immune pathway enrichment result. One possible explanation for this is that the GWAS loci identified in the new large PGC-AUT meta-analyses, and used here, differ substantially in their CpG methylation target locations compared to the initial PGC-AUT results that were derived using substantially fewer samples and used in our previous work. Instead, pathway enrichment analyses in this current study identified interesting tissue-specific pathways such as Wnt signaling in peripheral blood, which has been previously demonstrated to

be ASD-related via rare variant studies (Krumm et al., 2014). A placenta specific pathway targeted by ASD meQTLs was syncytiotrophoblast cell differentiation, an interesting finding given the fact that the syncytiotrophoblast layer of cells on placental villi is the only fetal tissue that makes contact with maternal blood (Gude, Roberts, Kalionis, & King, 2004). In other words, this result argues for an engagement of ASD SNPs in regulating maternal-fetal interactions during pregnancy, which are though to play a key role in fetal brain development (Hsiao & Patterson, 2012).

Consistent with our previous immune response findings, we observed enrichment in immune-response related pathways for fetal brain tissue (e.g. glial cell development) and virus-host interaction pathways in placenta. Placenta and fetal brain results are more likely to reflect causal mechanisms to ASD etiology given their time of collection relative to the gestational period (during for fetal brain and immediately following for placenta) and the frameworks connecting them to ASD. In contrast, cord blood and peripheral blood are mostly useful etiologically for their ability to serve as surrogates for less accessible tissues, though they may also reflect consequences of ASD (i.e instead of being causally related to ASD). The immune-response findings being specific to fetal brain and placenta then argue for their mechanistic, or causal role in ASD etiology. Future studies of ASD SNP controlled DNAm should acknowledge these respective roles for fetal brain and placenta vs. blood tissues, in order to aid interpretation of findings.

We identified putative downstream targets for the loci containing the 5 genome-wide significant SNPs associated with ASD in placenta and peripheral blood tissues for follow up functional analyses. Interestingly, *PINX1*, a new gene implicated by our mapping strategy that does not

contain the ASD-associated SNP itself, has been previously demonstrated to contain rare CNVs in ASD cases (Pinto et al., 2010). We did not identify any cord blood or fetal brain meQTLs in these genome-wide significant loci. One possible explanation for this fact is that the placenta and peripheral blood datasets offered the best opportunities to identify meQTLs, owing to a large preponderance of measured CpG sites via WGBS in placenta, and the large sample size using for peripheral blood meQTL detection. Therefore, incorporation of meQTLs derived from improved DNAm coverage and/or larger sample sizes may also identify cord blood and fetal brain meQTLs in these regions. According to the current evidence, these peripheral blood and placenta relationships do not reflect similar relationships in the brain and so instead offer insight into GWAS SNP functionality in these tissues themselves. In the case of the placenta, this may still offer causal insights, provided they fit the in framework linking altered placenta structure and function to an altered maternal-fetal interface, which itself leads to altered fetal brain development (Hsiao & Patterson, 2012).

Future work should continue to apply these approaches to integrate genetic and epigenetic data in an ASD context as both GWAS results improve and meQTL maps for different tissues become increasingly available to further aid in prioritization of potential functional targets of common genetic variants associated with ASD.

### 3.6 Conclusions

We derived a new comprehensive meQTL map in placenta tissue using unified genome-wide SNP and DNA methylation data from the same individuals. Using PGC-AUT GWAS results from the largest GWAS in autism to date, we replicate our previous finding of ASD-SNP enrichment for fetal brain and early-life peripheral blood meQTLs, and additionally show

enrichment for cord blood and placenta meQTLs. These findings shed light on ASD SNP mechanisms, specifically in their propensity to control DNAm across a range of tissue types, and argues for the continued investigation of the ASD-related methylation differences in these tissues for both mechanistic insights to and consequences of ASD. New tissue-specific and non-specific biological pathways, genes, and loci that are downstream targets of ASD risk SNPs enable prioritization of candidate genes to examine in future functional experiments to understand the mechanisms by which SNP variants influence autism disease risk.

### 3.7 References


1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., … McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., … Flicek, P. (2017). Ensembl 2017. *Nucleic Acids Research*, *45*(D1), D635–D642. https://doi.org/10.1093/nar/gkw1104

Aldinger, K. A., Plummer, J. T., & Levitt, P. (2013). Comparative DNA methylation among females with neurodevelopmental disorders and seizures identifies TAC1 as a MeCP2 target gene. *Journal of Neurodevelopmental Disorders*, *5*(1), 15. https://doi.org/10.1186/1866-1955-5-15

Andrews, S. V., Ellis, S. E., Bakulski, K. M., Sheppard, B., Croen, L. A., Hertz-Picciotto, I., … Fallin, M. D. (2017). Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. *Nature Communications*, *8*(1), 1011. https://doi.org/10.1038/s41467-017-00868-y

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism*, *8*, 21. https://doi.org/10.1186/s13229-017-0137-9

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., … Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology*, *12*(1), R10. https://doi.org/10.1186/gb-2011-12-1-r10

Berko, E. R., Suzuki, M., Beren, F., Lemetre, C., Alaimo, C. M., Calder, R. B., … Greally, J. M. (2014). Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genetics*, *10*(5), e1004402. https://doi.org/10.1371/journal.pgen.1004402

bsseq. (n.d.). Retrieved February 15, 2018, from http://bioconductor.org/packages/bsseq/

Budimirovic, D. B., & Kaufmann, W. E. (2011). What can we learn about autism from studying fragile X syndrome? *Developmental Neuroscience*, *33*(5), 379–394. https://doi.org/10.1159/000330213

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Centers for Disease Control and Prevention (CDC). (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years--Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., … Bolton, P. (2015). Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry*, *72*(5), 415–423. https://doi.org/10.1001/jamapsychiatry.2014.3028

Davis, L. K., Gamazon, E. R., Kistner-Griffin, E., Badner, J. A., Liu, C., Cook, E. H., … Cox, N. J. (2012). Loci nominally associated with autism from genome-wide analysis show enrichment of brain expression quantitative trait loci but not lymphoblastoid cell line expression quantitative trait loci. *Molecular Autism*, *3*(1), 3. https://doi.org/10.1186/2040-2392-3-3

Feinberg, J. I., Bakulski, K. M., Jaffe, A. E., Tryggvadottir, R., Brown, S. C., Goldman, L. R., … Feinberg, A. P. (2015). Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *International Journal of Epidemiology*, *44*(4), 1199–1210. https://doi.org/10.1093/ije/dyv028

Gamazon, E. R., Badner, J. A., Cheng, L., Zhang, C., Zhang, D., Cox, N. J., … Liu, C. (2013). Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Molecular Psychiatry*, *18*(3), 340–346. https://doi.org/10.1038/mp.2011.174

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Gregory, S. G., Connelly, J. J., Towers, A. J., Johnson, J., Biscocho, D., Markunas, C. A., … Pericak-Vance, M. A. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Medicine*, *7*, 62. https://doi.org/10.1186/1741-7015-7-62

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R., Won, H., … Børglum, A. D. (2017). Common risk variants identified in autism spectrum disorder. *bioRxiv*, 224774. https://doi.org/10.1101/224774

Gude, N. M., Roberts, C. T., Kalionis, B., & King, R. G. (2004). Growth and function of the normal human placenta. *Thrombosis Research*, *114*(5–6), 397–407. https://doi.org/10.1016/j.thromres.2004.06.038

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., … Mill, J. (2016). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nature Neuroscience*, *19*(1), 48–54. https://doi.org/10.1038/nn.4182

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, *13*(10), R83. https://doi.org/10.1186/gb-2012-13-10-r83

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., … Kent, W. J. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*, *34*(Database issue), D590-598. https://doi.org/10.1093/nar/gkj144

Hsiao, E. Y., & Patterson, P. H. (2012). Placental regulation of maternal-fetal interactions and brain development. *Developmental Neurobiology*, *72*(10), 1317–1326. https://doi.org/10.1002/dneu.22045

Ignatiadis, N., Klaus, B., Zaugg, J. B., & Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, *13*(7), 577–580. https://doi.org/10.1038/nmeth.3885

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216–221. https://doi.org/10.1038/nature13908

James, S. J., Shpyleva, S., Melnyk, S., Pavliv, O., & Pogribny, I. P. (2013). Complex epigenetic regulation of engrailed-2 (EN-2) homeobox gene in the autism cerebellum. *Translational Psychiatry*, *3*, e232. https://doi.org/10.1038/tp.2013.8

Jedele, K. B. (2007). The overlapping spectrum of rett and angelman syndromes: a clinical review. *Seminars in Pediatric Neurology*, *14*(3), 108–117. https://doi.org/10.1016/j.spen.2007.07.002

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, *32*(Database issue), D493-496. https://doi.org/10.1093/nar/gkh103

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences*, *37*(2), 95–105. https://doi.org/10.1016/j.tins.2013.11.005

Ladd-Acosta, C., Hansen, K. D., Briem, E., Fallin, M. D., Kaufmann, W. E., & Feinberg, A. P. (2014). Common DNA methylation alterations in multiple brain regions in autism. *Molecular Psychiatry*, *19*(8), 862–871. https://doi.org/10.1038/mp.2013.114

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., … Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

McClay, J. L., Shabalin, A. A., Dozmorov, M. G., Adkins, D. E., Kumar, G., Nerella, S., … van den Oord, E. J. C. G. (2015). High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biology*, *16*, 291. https://doi.org/10.1186/s13059-015-0842-7

missMethyl. (n.d.). Retrieved August 5, 2016, from http://bioconductor.org/packages/missMethyl/

Mount, R. H., Charman, T., Hastings, R. P., Reilly, S., & Cass, H. (2003). Features of autism in Rett syndrome and severe mental retardation. *Journal of Autism and Developmental Disorders*, *33*(4), 435–442.

Nagarajan, R. P., Hogart, A. R., Gwye, Y., Martin, M. R., & LaSalle, J. M. (2006). Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics*, *1*(4), e1-11.

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational Psychiatry*, *4*, e433. https://doi.org/10.1038/tp.2014.70

Newschaffer, C. J., Croen, L. A., Fallin, M. D., Hertz-Picciotto, I., Nguyen, D. V., Lee, N. L., … Shedd-Wise, K. M. (2012). Infant siblings and the investigation of autism risk factors. *Journal of Neurodevelopmental Disorders*, *4*(1), 7. https://doi.org/10.1186/1866-1955-4-7

Nguyen, A., Rauch, T. A., Pfeifer, G. P., & Hu, V. W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB*

*Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *24*(8), 3036–3051. https://doi.org/10.1096/fj.10-154484

O'Roak, B. J., Stessman, H. A., Boyle, E. A., Witherspoon, K. T., Martin, B., Lee, C., … Eichler, E. E. (2014). Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nature Communications*, *5*, 5595. https://doi.org/10.1038/ncomms6595

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., … Scherer, S. W. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, *94*(5), 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., … Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368–372. https://doi.org/10.1038/nature09146

PriorityPruner. (n.d.). Retrieved August 5, 2016, from http://prioritypruner.sourceforge.net/

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., … State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA*, *311*(17), 1770–1777. https://doi.org/10.1001/jama.2014.4144

Schroeder, D. I., Schmidt, R. J., Crary-Dooley, F. K., Walker, C. K., Ozonoff, S., Tancredi, D. J., … LaSalle, J. M. (2016). Placental methylome analysis from a prospective autism study. *Molecular Autism*, *7*, 51. https://doi.org/10.1186/s13229-016-0114-8

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics (Oxford, England)*, *28*(10), 1353–1358. https://doi.org/10.1093/bioinformatics/bts163

Shi, J., Marconett, C. N., Duan, J., Hyland, P. L., Li, P., Wang, Z., … Landi, M. T. (2014). Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nature Communications*, *5*, 3365. https://doi.org/10.1038/ncomms4365

Sun, W., Poschmann, J., Cruz-Herrera Del Rosario, R., Parikshak, N. N., Hajan, H. S., Kumar, V., … Prabhakar, S. (2016). Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell*, *167*(5), 1385–1397.e11. https://doi.org/10.1016/j.cell.2016.10.031

van Eijk, K. R., de Jong, S., Strengman, E., Buizer-Voskamp, J. E., Kahn, R. S., Boks, M. P., … Ophoff, R. A. (2015). Identification of schizophrenia-associated loci by combining DNA

methylation and gene expression data from whole blood. *European Journal of Human Genetics: EJHG*, *23*(8), 1106–1110. https://doi.org/10.1038/ejhg.2014.245

Volkov, P., Olsson, A. H., Gillberg, L., Jørgensen, S. W., Brøns, C., Eriksson, K.-F., … Ling, C. (2016). A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits. *PloS One*, *11*(6), e0157776. https://doi.org/10.1371/journal.pone.0157776

Wong, C. C. Y., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., … Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry*, *19*(4), 495–503. https://doi.org/10.1038/mp.2013.41

Zhu, L., Wang, X., Li, X.-L., Towers, A., Cao, X., Wang, P., … Jiang, Y.-H. (2014). Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. *Human Molecular Genetics*, *23*(6), 1563–1578. https://doi.org/10.1093/hmg/ddt547

**Table 3.1: meQTL enrichment statistics for ASD GWAS SNPs across 5 tissue types**

| | **ASD p-value = 1E-4** | | **ASD p-value = 1E-5** | |
|---|---|---|---|---|
| | ***meQTL p-value = 1E-8*** | | ***meQTL p-value = 1E-8*** | |
| **Fetal Brain[1]** | 3.94 (<0.001) | | 4.53 (<0.001) | |
| | ***meQTL p-value = 4.8E-10*** | | ***meQTL p-value = 4.8E-10*** | |
| **Adipose Tissue[1]** | 2.33 (< 0.001) | | 1.41 (0.317) | |
| | ***meQTL FDR = 5%*** | ***meQTL FDR = 1%*** | ***meQTL FDR = 5%*** | ***meQTL FDR = 1%*** |
| **Peripheral Blood[2]** | 1.30 (< 0.001) | 1.39 (< 0.001) | 1.37 (< 0.001) | 1.39 ( 0.011) |
| **Cord Blood[2]** | 2.00 (< 0.001) | 2.19 (< 0.001) | 1.98 (< 0.001) | 2.16 (< 0.001) |
| **Placenta[2]** | 1.14 (< 0.001) | 1.10 (0.011) | 1.23 (0.002) | 1.25 (0.011) |

Enrichment fold statistics and *P*-values based on 1,000 permutations. [1]LD pruning performed with 1000 Genomes CEU samples. [2]LD pruning performed with study-specific genotype data. See Methods for additional details.

**Table 3.2: Loci expansion results for AUT-PGC SNPs reaching genome-wide significance (p-value < 5E-8) in discovery sample and/or combined discovery and replication sample**

| SNP ID | Locus | Min SNP PGC p-value | Gene location of SNPs | Peripheral blood meQTL? | Cord blood meQTL? | Fetal brain meQTL? | Placenta meQTL? | Gene location of peripheral blood meQTL targets | Gene location of placenta meQTL targets | Figure depiction |
|---|---|---|---|---|---|---|---|---|---|---|
| rs910805 | chr20:21236264-21445443 | 2.04E-9 | XRN2, | ✓ | × | × | ✓ | NA, | KIZ, KIZ-AS1, XRN2, NKX2-4, | Fig. III-2 |
| rs10099100 | chr8:10547758-10576775 | 1.07E-8 | C8orf74, | ✓ | × | × | ✓ | SOX7, PINX1, | RP1L1, C8orf74, PINX1, | Fig. 3.2 |
| rs71190156 | chr20:14698357-14836243 | 2.75E-8 | MACROD2, | × | × | × | ✓ | NA, | MACROD2, | Fig. III-3 |
| rs111931861 | chr7:104744219-105064665 | 1.12E-7[a] | KMT2E, SRPK2, | × | × | × | ✓ | NA, | SRPK2, | Fig. III-4 |
| rs201910565 | chr1:96508040-96590987 | 3.41E-7[b] | OR4F16, OR4F3, OR4F29, RNU6-2, | × | × | × | ✓ | NA, | OR4F16, OR4F3, OR4F29, RNU6-2, | Fig. III-5 |

a: reached 3E-8 in combined discovery and replication sample

b: reached 4E-8 in combined discovery and replication sample

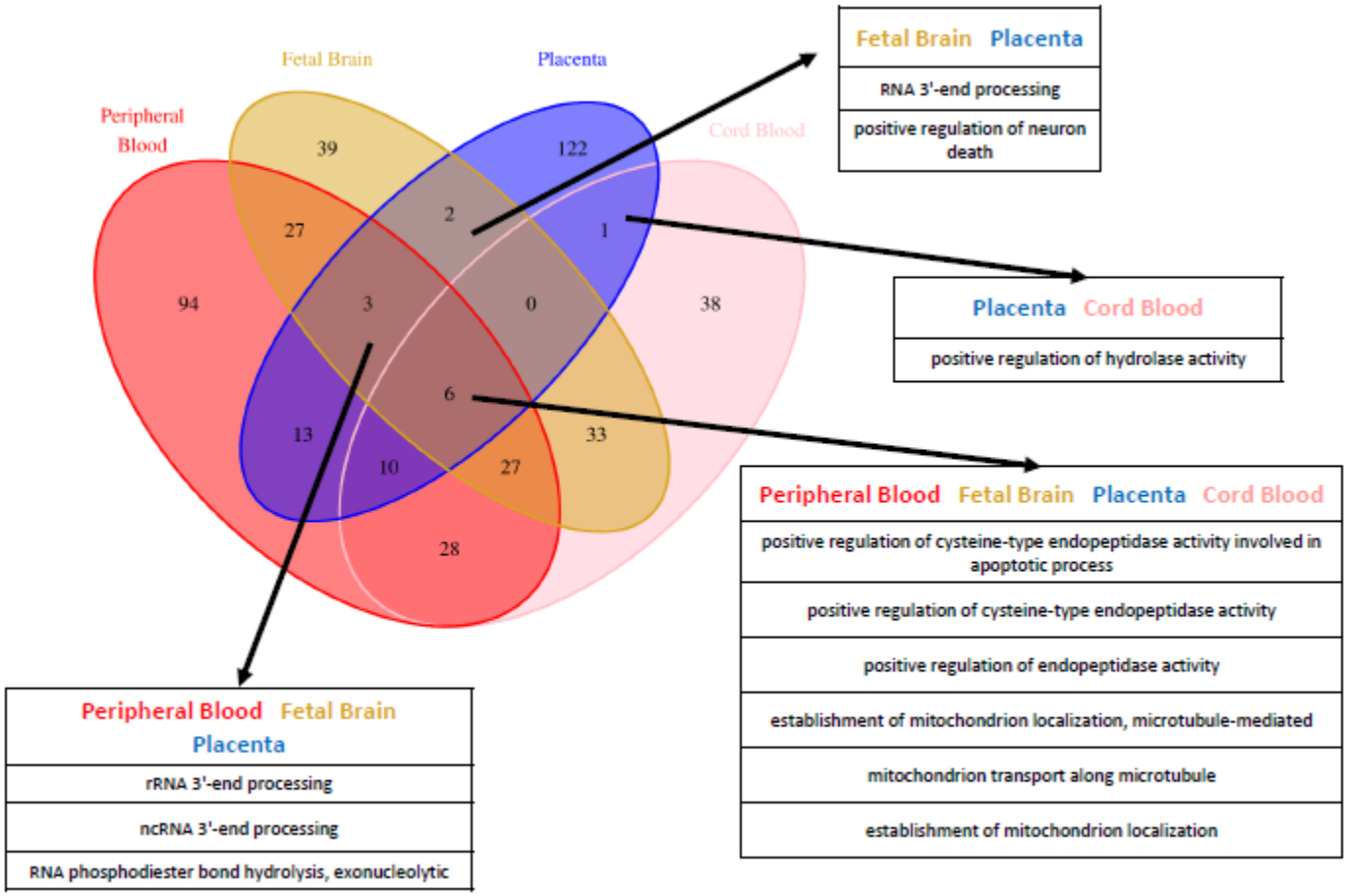



**Figure 3.1: meQTL target Gene ontology enrichment analysis**. Venn diagram depicts number of nominally significant ($p < 0.05$) biological process that overlap different tissue types.

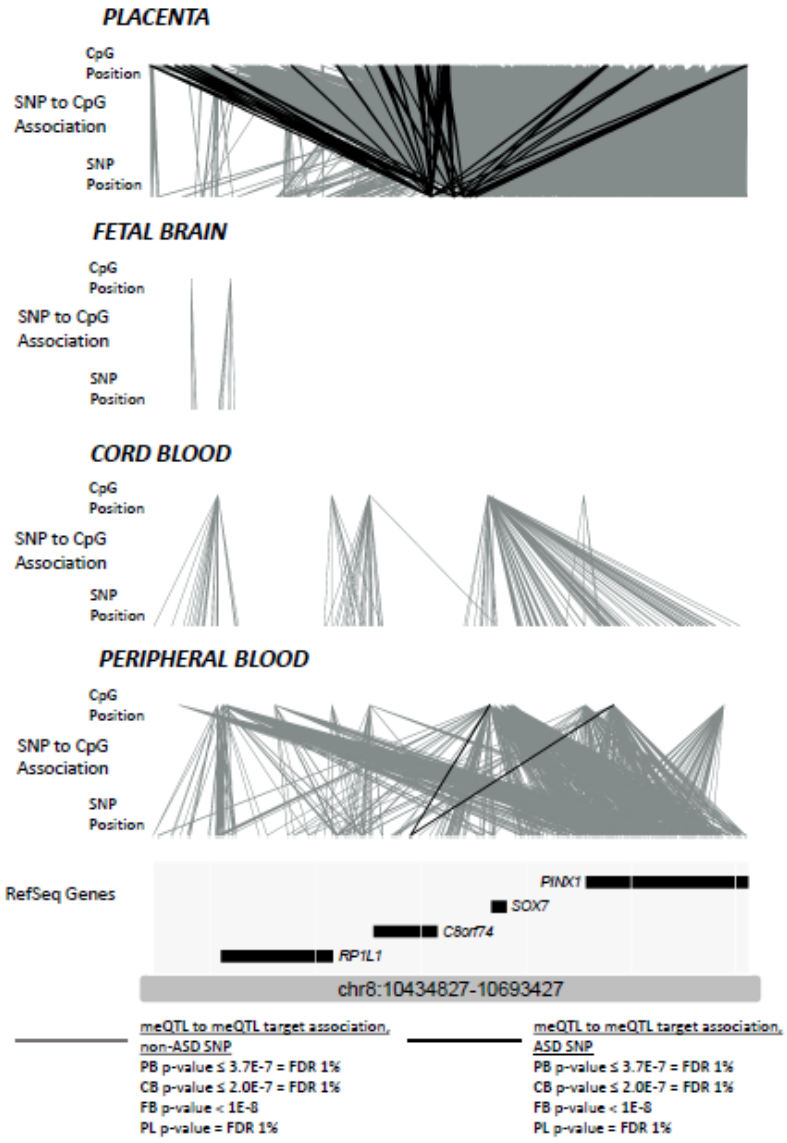



**Figure 3.2: Identification of novel candidate genes for ASD through meQTL mapping in peripheral blood, cord blood, fetal brain, and placenta for locus on chromosome 8.** Each tissue-specific panel presents, from bottom to top: genomic location, gene annotations, SNP locations, SNP-CpG associations, CpG locations. Light gray meQTL association lines denote all SNP to CpG associations in that tissue type. Dark meQTL association lines denote SNP-CpG associations for ASD-associated SNPs in PGC (*P*-value <= 1E-5). Data are presented for meQTL maps for (from top to bottom): placenta, fetal brain, cord blood, and peripheral blood. Please note locus coordinates differ from those in **Table 2.2** because in this context they encompass the locations of meQTL target CpG sites.

# CHAPTER 4: PLACENTA DNA METHYLATION AND AUTISM SPECTRUM DISORDER

## 4.1 Abstract

Background: Epigenetic mechanisms are of increasing interest to elucidate the etiology of autism spectrum disorder (ASD). Previous studies examining DNA methylation (DNAm), the most well studied epigenetic mark, and ASD have mostly utilized brain or blood tissue for DNAm measurement. Placenta DNAm is not well understood in an ASD context, despite the key role the placenta plays in regulating the in utero environment and the prenatal window being a critical time period for ASD risk factors.

Objectives: This study aims to investigate placenta methylation via comprehensive whole genome bisulfite sequencing (WGBS), which measures methylation at about 26 million loci, in order to identify A) differentially methylated regions (DMRs) and B) global methylation differences, related to ASD diagnosis. We conducted sex stratified analyses comparing the placentas from children who went on to be classified as having ASD ($N_{males}$ = 18) and who went on to be classified as non-typically developing (NTD; $N_{males}$ = 26, $N_{females}$ = 24) to those from children who went on to classified as typically developing (TD; $N_{males}$ = 17, $N_{females}$ = 20).

Results: No genome-wide significant ASD-associated DMRs were identified. However, NTD males displayed suggestive evidence for differences in global DNA methylation patterns in CpG island shores (P = 0.043) and shelves (P = 0.037), and in open sea locations (P = 0.014) relative to TD males. No significant global methylation differences between ASD and TD males or between NTD and TD females were observed.

Conclusions: No significant DMRs were identified in placentas obtained from women whose children went on to be diagnosed with ASD. Global methylation patterns did distinguish NTD males from TD males, but not to an extent that surpassed multiple testing correction. These results do not provide insight into the altered placental mechanisms that contribute to ASD etiology, nor do they provide placenta DNAm profiles that could serve as an ASD or NTD biomarker. Future studies in larger sample sizes will enable detection of autism-related differential methylation of smaller magnitude. These studies should be undertaken given the strong theoretical framework linking placenta DNAm to fetal brain developing, and the previous successful placenta DNAm studies in the context of ASD and related phenotypes.

## 4.2 Introduction

Autism spectrum disorder (ASD) is characterized by social, behavioral, and communication deficits and affects 1 in 68 children in the US (Christensen et al., 2016). ASD is known to have a strong genetic basis (Colvert et al., 2015; Gaugler et al., 2014; Sandin et al., 2014). Common variant findings are increasingly being realized through the Psychiatric Genomics Consortium ASD working group (PGC-AUT), with the first mega-analysis results recently being published (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017) and the next iteration of summary statistics recently becoming available (Grove et al., 2017). Most genetic findings to date have come from rare variant studies and include identification of inherited and *de novo* single nucleotide and copy number variants (CNVs) associated with ASD (Iossifov et al., 2014; O'Roak et al., 2014; Sanders et al., 2015).

Increasingly, epigenetic factors have been investigated in ASD because of multiple lines of suggesting its role in the disease process. First, epigenetic mechanisms have been implicated by

ASD rare variant findings (Krumm, O'Roak, Shendure, & Eichler, 2014; Pinto et al., 2014; Sanders et al., 2015). Second, many disorders (Rett, Angelman, Fragile X) that overlap with ASD phenotypically have an epigenetic basis (Budimirovic & Kaufmann, 2011; Jedele, 2007; Mount, Charman, Hastings, Reilly, & Cass, 2003).

Previous studies of ASD and genome-wide DNA methylation patterns, measured in postmortem brain tissue because it is thought to be the affected tissue type in ASD, have implicated multiple genes including *TSAPN32*/*C11orf21*, *ZFP57*, *PRRT1*, and *SDHAP3* (Ladd-Acosta et al., 2014; Nardone et al., 2014). Candidate gene approaches have found associations with *MeCP2*, *OXTR*, *EN-2*, and *SHANK3* (Gregory et al., 2009; James, Shpyleva, Melnyk, Pavliv, & Pogribny, 2013; Nagarajan, Hogart, Gwye, Martin, & LaSalle, 2006; Zhu et al., 2014). Another recent postmortem brain-based study interrogated another type of epigenetic mark, histone acetylation, on a genome-wide scale. The authors reported the discovery of large scale dysregulation of histone 3 lysine 27 acetylation (H3K27ac) in a subset of ASD cerebral cortex samples (Sun et al., 2016). Another previous study also identified various ASD-associated changes in histone 3 lysine 4 methylation (H3K4me3) comparing postmortem prefrontal cortex samples from ASD cases to controls (Shulha et al., 2012).

Because sample sizes for brain-based studies are limited and there could be confounding issues related to using postmortem samples, complementary blood-based epigenetic studies of ASD have been undertaken as well. These studies have implicated various genes such as *NFYC*, *PTPRCAP*, and *MBD4* (Aldinger, Plummer, & Levitt, 2013; Gregory et al., 2009; Wong et al., 2014), There have also been ASD-based investigations of DNAm from buccal epithelium (Berko et al., 2014), lymphoblastoid cell lines (Nguyen, Rauch, Pfeifer, & Hu, 2010), and the sperm of

fathers of children with ASD (Feinberg et al., 2015). Still, in an ASD context, the accessibility of DNAm from peripheral tissues and the subsequent potential for epidemiological scale cohorts comes at the cost of a lack of direct etiologic relevance of findings.

Given that epigenetic marks are tissue-specific, the choice of source tissue for epigenetic association analysis needs to be fully considered (Bakulski, Halladay, Hu, Mill, & Fallin, 2016). A tissue that balances etiologic relevance in an ASD context and accessibility is the placenta, which functions as unique maternal-fetal interface during pregnancy. The current framework linking placenta function to increased susceptibility to neurodevelopmental disorders involves the placenta's role in regulating maternal-fetal interactions. Overall, maternal insults such as immune activation, malnutrition, or stress are thought to lead to altered placenta status in the state of cytokines, hormones and endocrine factors, and gene expression. These altered states then impact placental processes such as immune tolerance, the growth of the placenta into the uterine wall, and the production of neuroactive molecules. Finally, these affected processes result in several adverse events linked to the prenatal programming of disease, such as immune dysfunction, intrauterine growth restriction or malnutrition, hypoxia, ischemia, and altered neuronal development and circuit formation, each of which contribute to an increased susceptibility to neurodevelopmental disorders (Hsiao & Patterson, 2012). Moreover, there are differences in placental gene expression even in normal development between male and female fetuses (Buckberry, Bianco-Miotto, Bent, Dekker, & Roberts, 2014; Clifton, 2010; Cvitic et al., 2013), a fact that is particularly intriguing with respect to ASD given its high male to female ratio in prevalence (Christensen et al., 2016).

In keeping with this framework, several studies have identified links between altered placenta structure and/or function and ASD. One study based in the prospective, enriched risk birth cohort Markers of Autism Risk in Babies – Learning Early Signs (MARBLES), demonstrated that mothers with a previous child with diagnosed ASD had an 8-fold greater odds of having placentas (from a subsequent pregnancy) with two or more trophoblast inclusions compared to control mothers (Walker et al., 2013). Trophoblast inclusions are a hallmark of abnormal placental growth, and are characterized by the typically inner cytotrophoblast cell layer of placental villi enveloping the typically outer syncytiotrophoblast layer (Adler, Madankumar, Rosner, & Reznik, 2014). In addition to this study examining pregnancies subsequent to the ASD-affected child, similar relationships demonstrating an association between trophoblast inclusions and ASD have also come in directly comparing placentas from children who went on to develop ASD to those from control children (Anderson, Jacobs-Stannard, Chawarska, Volkmar, & Kliman, 2007). Another more recent study also linked various aspects of placenta histopathology to ASD. Specifically, acute placenta inflammation, chronic uteroplacental vasculitis, the fetal inflammatory response in the chorionic plate, and maternal vascular malperfusion pathology were all associated with placentas from children who went on to develop ASD compared to those from matched controls (Straughen et al., 2017). Finally, another paper demonstrated the utility of random forest-based classification of placenta chorionic surface vascular network structure to distinguish between placentas from high risk ASD pregnancies (i.e. those from mothers with a child previously diagnosed with ASD) and population-based controls (Chang et al., 2017).

Understanding the epigenetic features dysregulated in ASD could shed increased light on the placenta structure and/or functional mechanisms that impair fetal brain development. Moreover, placenta-based signatures of ASD could serve as a useful early biomarker for ASD that could guide early interventions, which have been shown to improve prognosis (Klintwall, Eldevik, & Eikeseth, 2015). With respect to placenta epigenetic features specifically and neurodevelopment, there has been a wave of studies using both candidate gene and genome-wide approaches to link infant-based neurobehavioral outcomes and DNAm (Lester & Marsit, 2018). There has been one previous study specifically examining DNAm profiles and the ASD phenotype in placenta. This study identified one autosomal differentially methylated region in *DLL1* (Schroeder et al., 2016), near a putative fetal brain enhancer. This study was particularly distinguished by its used of sequencing-based DNAm, which allows for greater methylome coverage than Illumina array platforms typically used for epigenome-wide association studies.

Our study improves upon this previous placenta DNAm and ASD study by also using placenta samples measured for DNAm using sequencing, but via a comprehensive technique known as whole genome bisulfite sequencing (WGBS) as opposed to the MethylC-seq technique used previously. In addition, our study makes use of a greater number of samples overall (N = 106 vs N = 47), and includes samples with an additional, mutually exclusive phenotype category known as non-typical development (NTD), allowing for a more nuanced investigation of the relationship between placenta DNAm and neurodevelopment more broadly. Using placental samples from an enriched risk birth cohort, we searched for differentially methylated regions and global methylation differences in placenta tissue comparing pregnancies where the children went on to be diagnosed with ASD, or children who were diagnosed with non-typical but not ASD

development (NTD), to placentas from children who went on to be classified as typically developing (TD). We perform these analyses in a sex-stratified framework to acknowledge the known sexual dimorphism in placenta gene expression (Buckberry et al., 2014; Cvitic et al., 2013) and DNAm (Gong et al., 2018; Martin et al., 2017).

## 4.3 Methods

### *4.3.1 EARLI Study Description*

The Early Autism Risk Longitudinal Investigation (EARLI) is a prospective, enriched-risk birth cohort that has been described in detail elsewhere (Newschaffer et al., 2012). The EARLI study was approved by Human Subjects Institutional Review Boards (IRBs) from each of the four study sites (Johns Hopkins University, Drexel University, University of California Davis, and Kaiser Permanente). Informed consent was obtained from all participating families. The 232 mothers with a subsequent child born through this study had births between November 2009 and March 2012. Placental biopsy samples were collected after delivery at each clinical lab site using Baby Tischler Punch Biopsy Forceps. Samples were stored at ambient temperature in RNAlater vials (Qiagen, Cat. No. 76154) and shipped same-day to the Johns Hopkins Biological Repository (JHBR) in Baltimore, Maryland, for storage at -190°C until processing. After biopsy, the placentas were then placed flat in a bag with 10% formalin at a volume roughly equivalent to the size of the placenta.

Infants were followed with extensive neurophenotyping until age three, when they obtained an ASD diagnosis. Children were classified as having autism spectrum disorder (ASD) if they met or exceeded the ASD cutoff of the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) and met the Diagnostic and Statistical Manual of Mental Disorders, 4$^{th}$ edition, Text

Revision (DSM-IV-TR) ("DSM-IV: Diagnostic and Statistical Manual of Mental Disorders | JAMA | The JAMA Network," n.d.) criteria for Autistic Disorder or Pervasive Developmental Disorder – Not otherwise specified (PDD-NOS). Children were classified as having non-typical development (NTD) if they did not meet the criteria for ASD classification but had two or more Mullen Scales of Early Learning (Mullen, 1995) subtests ≥ 1.5 SD below the mean and/or one or more Mullen subtests ≥ 2 SD below the mean and/or ADOS ≤ 3 points below the ASD cutoff. Children were classified as having typical development (TD) if they did not meet the criteria for ASD classification and had no more than one Mullen subtest ≥ 1.5 SD below the mean and no Mullen subtest ≥ 2 SD below the mean and ADOS > 3 points below the ASD cutoff.

*4.3.2 Whole genome bisulfite sequencing, alignment, and quality control*

Genomic DNA (gDNA) from biopsy punches was extracted at JHBR using a QIAgen QIAsymphony automated workstation with the Blood 1000 protocol of the DSP DNA Midi kit (Cat. No. 937255) as per manufacturer's instructions. 250ng gDNA was bisulfite treated using the Zymo EZ-96 DNA Methylation Kit (Cat. No. D5004), and sequencing libraries were made with the NEBNext Ultra DNA Library Prep Kit for Illumina. Library preparation was performed at the Center for Epigenetics and libraries were sequenced at ~4x coverage at the High Throughput Sequencing Center at the Johns Hopkins University School of Medicine using the 3 Illumina HiSeq 2500 instruments. These low coverage measurements were ran in order to maximize sample numbers, exploiting known smoothing techniques to recapture higher coverage sequencing-based DNAm from low coverage data (Hansen, Langmead, & Irizarry, 2012). Initially 133 placenta samples were run on 16 flow cells; two samples failed due to a fluidics error and were re-run using a 17$^{th}$ flow cell.

Raw .fastq files were aligned the human reference genome (hg38) downloaded from Ensembl (Aken et al., 2017) via Bismark (v0.16.3) (Krueger & Andrews, 2011) using the default arguments for paired-end sequencing data. M-bias plots (Hansen et al., 2012) were used to inform the trimming of the 5bp of both the first and second reads, and methylation status was called using the Bismark methylation extractor function. All subsequent analyses used R (version 3.4.2) and the R package *bsseq* (version 1.13.6) ("bsseq," n.d.). The bsseq function *read.bismark()* was used to read in Bismark output files ("cytosine report" format) and obtain DNAm measurements on a per-CpG site basis. A total of 29,091,077 CpG sites were measured with at least 1 read in 1 sample out of the total 133. Next methylation levels were smoothed using the BSmooth algorithm that has been previously described (Hansen et al., 2012). Briefly, this technique employs a local-likelihood smoothing algorithm to recapitulate higher coverage whole genome bisulfite sequencing data from lower coverage data, exploiting known local correlation structure in DNAm data. We used the *BSmooth()* function in the *bsseq* R package with the default arguments.

### *4.3.3 Identification of differentially methylated regions*

Given the known sexual dimorphism of the placenta (Buckberry et al., 2014; Clifton, 2010; Cvitic et al., 2013; Gong et al., 2018; Martin et al., 2017), as well as the established sex-biased ratio of ASD diagnosis (Christensen et al., 2016), we elected to perform sex-stratified queries related to diagnosis. For males, we compared ASD cases to TD individuals, as well as NTD individuals to TD. For females, we only compared NTD to TD, because just a single female sample with available DNAm data was diagnosed with ASD. For each analysis, we limited DNAm based on sufficient coverage as recommended in *bsseq*. For the ASD vs. TD analysis in

males, we retained loci with ≥ 2 reads in at least 10 ASD samples and 10 TD samples, leaving a total of 26,988,836 DNAm sites across the 35 samples. For the NTD vs TD analysis in males, we retained loci with ≥ 2 reads in at least 15 NTD samples and 10 TD samples, leaving a total of 27,010,299 DNAm sites across the 43 samples. For the NTD vs TD analysis in females, we retained loci with ≥ 2 reads in at least 15 NTD samples and 10 TD samples as well, leaving a total of 26,988,807 DNAm sites across the 44 samples. Finally, we used the *BSmooth.tstat.fix()* and *dmrFinder()* functions in *bsseq* with the default arguments to identify DMRs. Statistical significance was assessed by creating 1000 null sets of methylation data via scrambling sample labels for diagnosis and re-running the DMR-finding algorithm; p-values were calculated as the number of null sets in which at least 1 DMR had a ≥ 'areaStat' value (sum of t-statistics) and width (in bp) than the candidate DMR.

*4.3.4 Global methylation analysis*

In addition to searching for DMRs, we also sought to compare global methylation levels between each of our 3 outcome groups (ASD, NTD, and TD). To do this, we first employed a similar coverage-based filtering of CpG sites as performed prior to the DMR queries for the full dataset of 29,091,077 CpG sites and 133 samples. Specifically we limited the data to retain sites with ≥ 2 reads in more than two-thirds of the total number of samples, leaving 26,752,597 DNAm sites. Next we downloaded a list of CpG islands from the UCSC Genome Browser Table (Karolchik et al., 2004), and defined CpG island shores as being 2000 bp upstream and downstream of these islands, and defined CpG island shelves as being 2000 bp flanking those shores. We used the function *findOverlaps()* in the *GenomicRanges* R package to classify DNAm sites into these 3 regions. Sites that did not overlap with any of these regions were classified as 'Open Sea' sites.

Next we ran the *clusterMaker()* function in the *bsseq* R package to collapse loci into regions in an identical manner as to how they are combined for DMR finding. We computed the mean within each cluster for each sample. If a cluster encompassed more than one genomic region (such as, for example, and island and shore in the same identified cluster), we computed the mean within each of the genomic region. We finally collapsed across clusters, ultimately retaining a mean for each sample within each of the 4 genomic regions. We compared these values in the same stratified groups as for the DMR analysis via t-tests.

## 4.4 Results

### *4.4.1 Identification of specific genomic locations showing methylation differences related to ASD or non-typical development*

Fetal-side placenta DNAm measurement at approximately 27 million CpG sites was available for 133 samples, of which 106 had an available diagnosis (**Supplementary Table IV-1**) at age 3. We conducted male-stratified analyses to identify DMRs associated with ASD (N = 18) or non-typically developing (N = 26) outcomes compared to typically developing male (N = 17) children; no genome-wide significant differences were observed **(Supplementary Figures IV-1** to **IV-6**). Similarly, female-stratified analyses revealed no genome-wide significant differences in locus-specific methylation related to non-typically developing (N = 24) compared to typically developing females (N = 20) (**Supplementary Figures IV-7** to **IV-9**).

### *4.4.2 Global methylation analyses*

We next conducted sex-stratified analyses comparing overall summed mean methylation levels between ASD, NTD, and TD diagnostic groups in various genomic contexts including CpG shores, shelves, islands, and open seas. We observed suggestive evidence that non-typically developing males had more hypomethylation compared to typically developing males in CpG

island shores (P = 0.043), CpG island shelves (P = 0.037), and in open sea locations across the genome (P = 0.014; **Figure 4.1 and Supplementary Table IV-2**). However, these differences were not significant after employing multiple testing correction ($p_{Bonferroni}$ = 0.05/12 tests = 0.0042). Moreover, effect sizes distinguishing the NTD and TD groups were small, and global methylation distributions were very similar. For example, the mean methylation level for NTD male samples in open sea locations, globally, was 0.608, which was only slightly different from the open sea methylation level (0.618) in TD males (**Supplementary Table IV-2**). Global methylation levels in ASD males were not significantly different from TD males in any genomic context, nor were NTD and TD female methylation levels (**Figure 4.1**).

### 4.5 Discussion

In this study, we sought to identify differences in placenta DNA methylation related to altered neurodevelopment. We applied two analytic approaches. First, we performed analyses to identify specific genomic regions showing differences in methylation between autistic or non-typically developing children relative to those with typical development. Second, we examined global measures of methylation at pre-defined genomic regions to identify types of genomic features associated with autism or non-typical neurodevelopment. We did not identify any DMRs or global methylation differences that were significant after multiple testing correction.

Despite many previous studies linking placenta DNAm to infant neurobehavior (Lester & Marsit, 2018) and one study to ASD (Schroeder et al., 2016), we did not identify any genome-wide significant DMRs separating ASD and ASD-related diagnostic groups from TD children. We chose to employ sex-stratified association analyses in light of the strong ASD male-to-female-bias and known sexual dimorphism of placenta DNA and gene expression. However, this

approach resulted in sample sizes of at most 26 samples per diagnostic outcome (TD male group). Thus, even though we have one of the largest placenta WGBS sample sizes, we may lack the statistical power to detect DMRs at effect sizes in placenta relevant to ASD. Differences between our analytic approach and a previous study of placenta DNAm and ASD that identified a methylation difference using a similar sample size (n=24 ASD and 23 TD kids) in a similar enriched birth cohort as EARLI may explain the different findings. For example, the previous study did not perform sex-stratified analyses and also did not test explicitly for DMRs but rather for average methylation differences between ASD and TD samples in both partially methylated domains (PMDs) and highly-methylated domains, or HMDs (Schroeder et al., 2016).

Additionally, we did not employ any direct correction for cell type heterogeneity given the lack of a current placenta cell type DNAm reference panel as has been developed for a variety of other tissues (Bakulski, Feinberg, et al., 2016; Guintivano, Aryee, & Kaminsky, 2013; Reinius et al., 2012; Teschendorff et al., 2016), and the lack of many methodological approaches for reference-free based correction in sequencing-based DNAm (Teschendorff & Zheng, 2017). Both of these limitations represent key areas of future research that should be pursued. Though customarily thought to be critical to avoid Type I error, lack of adjustment for cell type heterogeneity can result in decreased power in association studies as well (Teschendorff & Zheng, 2017), and may have impaired our ability to detect genome-wide significant DMRs.

We did identify suggestive global methylation differences between NTD and TD samples in males only. While our DMR analysis chiefly aims to identify regions of 1-2 Kb in length, this approach tests DNAm over much larger regions (for examples, all clusters contained in CpG islands, see **Methods**), albeit collapsing measurements into a single value per sample, per region.

In this manner, this analysis was more consistent with the previous ASD-placenta DNAm study that examined PMDs and HMDs that are, on average, an order of magnitude larger than these sought-after DMRs (Schroeder et al., 2013). However, suggestive differences in global methylation patterns between NTD and TD males did not surpass multiple testing correction and were characterized by very small effect sizes as far as mean differences between groups and overall differences of distributions. Moreover, these findings are limited somewhat by the lack of a "dose-response" type relationship in ASD samples, such as a stronger degree of hypomethylation relative to TD samples in the more severe phenotype of ASD. Instead, we often observed an opposite effect size in the ASD to TD comparison to what we saw in the NTD to TD comparison. Still, it is theoretically possible that these hypomethylation effects are specific to the NTD diagnosis and this finding should be investigated further.

Future analyses of placenta DNAm and ASD should continue to pursue measurement of DNAm via WGBS, despite lack of significant findings in this study. Additional WGBS samples can be combined with the samples used in this study to better exploit the unbiased assessment of the placenta DNA methylome. Additionally, better leveraging of the PMD/HMD distinction can be made for association analyses. Specifically, it has been observed that greater inter-individual variability is present in placenta PMDs compared to HMDs (Schroeder et al., 2013, 2016). Therefore, placental DNAm association analyses, especially those of limited sample sizes needing to overcome multiple testing burden from a large number of interrogated loci presented by WGBS, can limit initial discovery regions to PMDs, as their greater variability dictates an increased likelihood to display phenotypic-related differences. However, this approach does not fully exploit the comprehensive coverage of the methylome offered by WGBS, and comes

despite the fact that the one significant difference observed between ASD and TD samples in the previous placenta DNAm-ASD study was in a HMD (Schroeder et al., 2016). Finally, future studies should also look to more explicitly incorporate exposure information during pregnancy and maternal characteristics to discovery placenta DNAm signatures of ASD. For example, it is possible that placenta DNAm patterns differ according to ASD diagnosis only in the presence of particular environmental exposures that affect the maternal-fetal interface. In this manner, placenta DNAm offers interesting potential to discover the joint effect of genetic and environmental risk factors on ASD.

### 4.6 Conclusions

This study fails to identify significant DMRs or global methylation differences in placentas from children later diagnosed with ASD or NTD from those with TD. Given the strong theoretical framework linking placenta DNAm, placenta structure and function, and altered fetal brain development, as well as previous studies successfully identifying placenta DNAm signatures of infant neurobehavior and ASD, additional studies of placenta DNAm and ASD are still warranted, albeit with larger sample sizes than those used in this study.

### 4.7 References

Adler, E., Madankumar, R., Rosner, M., & Reznik, S. E. (2014). Increased placental trophoblast inclusions in placenta accreta. *Placenta*, *35*(12), 1075–1078. https://doi.org/10.1016/j.placenta.2014.09.014

Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., … Flicek, P. (2017). Ensembl 2017. *Nucleic Acids Research*, *45*(D1), D635–D642. https://doi.org/10.1093/nar/gkw1104

Aldinger, K. A., Plummer, J. T., & Levitt, P. (2013). Comparative DNA methylation among females with neurodevelopmental disorders and seizures identifies TAC1 as a MeCP2 target

gene. *Journal of Neurodevelopmental Disorders*, *5*(1), 15. https://doi.org/10.1186/1866-1955-5-15

Anderson, G. M., Jacobs-Stannard, A., Chawarska, K., Volkmar, F. R., & Kliman, H. J. (2007). Placental trophoblast inclusions in autism spectrum disorder. *Biological Psychiatry*, *61*(4), 487–491. https://doi.org/10.1016/j.biopsych.2006.03.068

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism*, *8*, 21. https://doi.org/10.1186/s13229-017-0137-9

Bakulski, K. M., Feinberg, J. I., Andrews, S. V., Yang, J., Brown, S., L McKenney, S., … Fallin, M. D. (2016). DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics*, *11*(5), 354–362. https://doi.org/10.1080/15592294.2016.1161875

Bakulski, K. M., Halladay, A., Hu, V. W., Mill, J., & Fallin, M. D. (2016). Epigenetic Research in Neuropsychiatric Disorders: the "Tissue Issue." *Current Behavioral Neuroscience Reports*, *3*(3), 264–274. https://doi.org/10.1007/s40473-016-0083-4

Berko, E. R., Suzuki, M., Beren, F., Lemetre, C., Alaimo, C. M., Calder, R. B., … Greally, J. M. (2014). Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genetics*, *10*(5), e1004402. https://doi.org/10.1371/journal.pgen.1004402

bsseq. (n.d.). Retrieved February 15, 2018, from http://bioconductor.org/packages/bsseq/

Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A., & Roberts, C. T. (2014). Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. *Molecular Human Reproduction*, *20*(8), 810–819. https://doi.org/10.1093/molehr/gau035

Budimirovic, D. B., & Kaufmann, W. E. (2011). What can we learn about autism from studying fragile X syndrome? *Developmental Neuroscience*, *33*(5), 379–394. https://doi.org/10.1159/000330213

Chang, J.-M., Zeng, H., Han, R., Chang, Y.-M., Shah, R., Salafia, C. M., … Croen, L. (2017). Autism risk classification using placental chorionic surface vascular network features. *BMC Medical Informatics and Decision Making*, *17*(1), 162. https://doi.org/10.1186/s12911-017-0564-8

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Centers for Disease Control and Prevention (CDC). (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years--Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Clifton, V. L. (2010). Review: Sex and the human placenta: mediating differential strategies of fetal growth and survival. *Placenta*, *31 Suppl*, S33-39. https://doi.org/10.1016/j.placenta.2009.11.010

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., … Bolton, P. (2015). Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry*, *72*(5), 415–423. https://doi.org/10.1001/jamapsychiatry.2014.3028

Cvitic, S., Longtine, M. S., Hackl, H., Wagner, K., Nelson, M. D., Desoye, G., & Hiden, U. (2013). The human placental sexome differs between trophoblast epithelium and villous vessel endothelium. *PloS One*, *8*(10), e79233. https://doi.org/10.1371/journal.pone.0079233

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders | JAMA | The JAMA Network. (n.d.). Retrieved February 15, 2018, from https://jamanetwork.com/journals/jama/article-abstract/379036?redirect=true

Feinberg, J. I., Bakulski, K. M., Jaffe, A. E., Tryggvadottir, R., Brown, S. C., Goldman, L. R., … Feinberg, A. P. (2015). Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *International Journal of Epidemiology*, *44*(4), 1199–1210. https://doi.org/10.1093/ije/dyv028

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Gong, S., Johnson, M. D., Dopierala, J., Gaccioli, F., Sovio, U., Constância, M., … Charnock-Jones, D. S. (2018). Genome-wide oxidative bisulfite sequencing identifies sex-specific methylation differences in the human placenta. *Epigenetics*, 1–32. https://doi.org/10.1080/15592294.2018.1429857

Gregory, S. G., Connelly, J. J., Towers, A. J., Johnson, J., Biscocho, D., Markunas, C. A., … Pericak-Vance, M. A. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Medicine*, *7*, 62. https://doi.org/10.1186/1741-7015-7-62

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R., Won, H., … Børglum, A. D. (2017). Common risk variants identified in autism spectrum disorder. *bioRxiv*, 224774. https://doi.org/10.1101/224774

Guintivano, J., Aryee, M. J., & Kaminsky, Z. A. (2013). A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics*, *8*(3), 290–302. https://doi.org/10.4161/epi.23924

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, *13*(10), R83. https://doi.org/10.1186/gb-2012-13-10-r83

Hsiao, E. Y., & Patterson, P. H. (2012). Placental regulation of maternal-fetal interactions and brain development. *Developmental Neurobiology*, *72*(10), 1317–1326. https://doi.org/10.1002/dneu.22045

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216–221. https://doi.org/10.1038/nature13908

James, S. J., Shpyleva, S., Melnyk, S., Pavliv, O., & Pogribny, I. P. (2013). Complex epigenetic regulation of engrailed-2 (EN-2) homeobox gene in the autism cerebellum. *Translational Psychiatry*, *3*, e232. https://doi.org/10.1038/tp.2013.8

Jedele, K. B. (2007). The overlapping spectrum of rett and angelman syndromes: a clinical review. *Seminars in Pediatric Neurology*, *14*(3), 108–117. https://doi.org/10.1016/j.spen.2007.07.002

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, *32*(Database issue), D493-496. https://doi.org/10.1093/nar/gkh103

Klintwall, L., Eldevik, S., & Eikeseth, S. (2015). Narrowing the gap: effects of intervention on developmental trajectories in autism. *Autism: The International Journal of Research and Practice*, *19*(1), 53–63. https://doi.org/10.1177/1362361313510067

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences*, *37*(2), 95–105. https://doi.org/10.1016/j.tins.2013.11.005

Ladd-Acosta, C., Hansen, K. D., Briem, E., Fallin, M. D., Kaufmann, W. E., & Feinberg, A. P. (2014). Common DNA methylation alterations in multiple brain regions in autism. *Molecular Psychiatry*, *19*(8), 862–871. https://doi.org/10.1038/mp.2013.114

Lester, B. M., & Marsit, C. J. (2018). Epigenetic mechanisms in the placenta related to infant neurodevelopment. *Epigenomics*. https://doi.org/10.2217/epi-2016-0171

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Martin, E., Smeester, L., Bommarito, P. A., Grace, M. R., Boggess, K., Kuban, K., … Fry, R. C. (2017). Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period. *Epigenomics*, *9*(3), 267–278. https://doi.org/10.2217/epi-2016-0132

Mount, R. H., Charman, T., Hastings, R. P., Reilly, S., & Cass, H. (2003). Features of autism in Rett syndrome and severe mental retardation. *Journal of Autism and Developmental Disorders*, *33*(4), 435–442.

Mullen, E. M. (1995). Mullen Scales of Early Learning. *American Guidance Service Inc.*

Nagarajan, R. P., Hogart, A. R., Gwye, Y., Martin, M. R., & LaSalle, J. M. (2006). Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics*, *1*(4), e1-11.

Nardone, S., Sams, D. S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M., & Elliott, E. (2014). DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational Psychiatry*, *4*, e433. https://doi.org/10.1038/tp.2014.70

Newschaffer, C. J., Croen, L. A., Fallin, M. D., Hertz-Picciotto, I., Nguyen, D. V., Lee, N. L., … Shedd-Wise, K. M. (2012). Infant siblings and the investigation of autism risk factors. *Journal of Neurodevelopmental Disorders*, *4*(1), 7. https://doi.org/10.1186/1866-1955-4-7

Nguyen, A., Rauch, T. A., Pfeifer, G. P., & Hu, V. W. (2010). Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *24*(8), 3036–3051. https://doi.org/10.1096/fj.10-154484

O'Roak, B. J., Stessman, H. A., Boyle, E. A., Witherspoon, K. T., Martin, B., Lee, C., … Eichler, E. E. (2014). Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nature Communications*, *5*, 5595. https://doi.org/10.1038/ncomms6595

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., … Scherer, S. W. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, *94*(5), 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018

Reinius, L. E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S.-E., Greco, D., … Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PloS One*, *7*(7), e41361. https://doi.org/10.1371/journal.pone.0041361

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., … State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA*, *311*(17), 1770–1777. https://doi.org/10.1001/jama.2014.4144

Schroeder, D. I., Blair, J. D., Lott, P., Yu, H. O. K., Hong, D., Crary, F., … LaSalle, J. M. (2013). The human placenta methylome. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 6037–6042. https://doi.org/10.1073/pnas.1215145110

Schroeder, D. I., Schmidt, R. J., Crary-Dooley, F. K., Walker, C. K., Ozonoff, S., Tancredi, D. J., … LaSalle, J. M. (2016). Placental methylome analysis from a prospective autism study. *Molecular Autism*, *7*, 51. https://doi.org/10.1186/s13229-016-0114-8

Shulha, H. P., Cheung, I., Whittle, C., Wang, J., Virgil, D., Lin, C. L., … Weng, Z. (2012). Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. *Archives of General Psychiatry*, *69*(3), 314–324. https://doi.org/10.1001/archgenpsychiatry.2011.151

Straughen, J. K., Misra, D. P., Divine, G., Shah, R., Perez, G., VanHorn, S., … Salafia, C. M. (2017). The association between placental histopathology and autism spectrum disorder. *Placenta*, *57*, 183–188. https://doi.org/10.1016/j.placenta.2017.07.006

Sun, W., Poschmann, J., Cruz-Herrera Del Rosario, R., Parikshak, N. N., Hajan, H. S., Kumar, V., … Prabhakar, S. (2016). Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell*, *167*(5), 1385–1397.e11. https://doi.org/10.1016/j.cell.2016.10.031

Teschendorff, A. E., Gao, Y., Jones, A., Ruebner, M., Beckmann, M. W., Wachter, D. L., … Widschwendter, M. (2016). DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nature Communications*, *7*, 10478. https://doi.org/10.1038/ncomms10478

Teschendorff, A. E., & Zheng, S. C. (2017). Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics*, *9*(5), 757–768. https://doi.org/10.2217/epi-2016-0153

Walker, C. K., Anderson, K. W., Milano, K. M., Ye, S., Tancredi, D. J., Pessah, I. N., … Kliman, H. J. (2013). Trophoblast inclusions are significantly increased in the placentas of children in families at risk for autism. *Biological Psychiatry*, *74*(3), 204–211. https://doi.org/10.1016/j.biopsych.2013.03.006

Wong, C. C. Y., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., … Mill, J. (2014). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry*, *19*(4), 495–503. https://doi.org/10.1038/mp.2013.41

Zhu, L., Wang, X., Li, X.-L., Towers, A., Cao, X., Wang, P., … Jiang, Y.-H. (2014). Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. *Human Molecular Genetics*, *23*(6), 1563–1578. https://doi.org/10.1093/hmg/ddt547

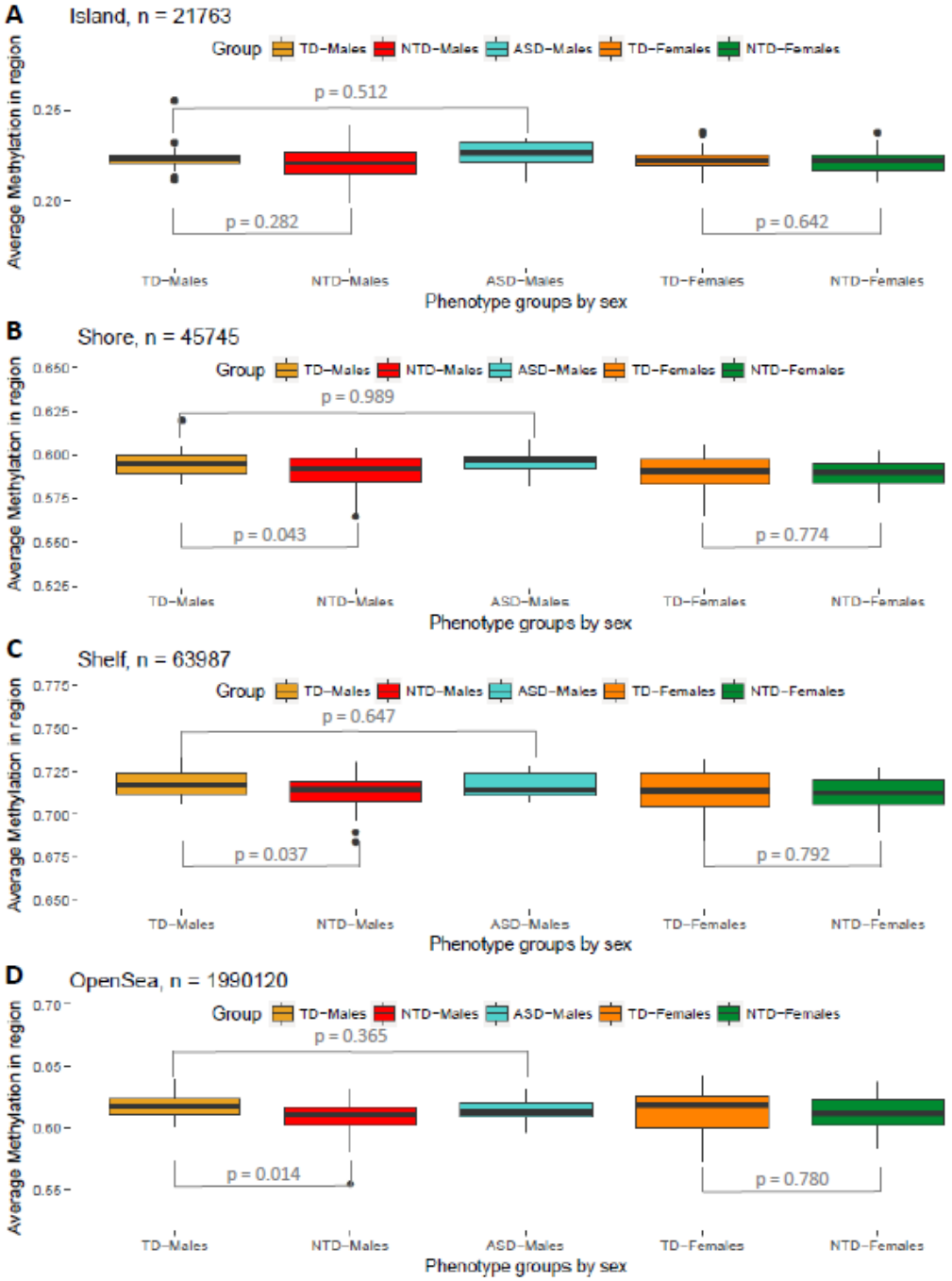



**Figure 4.1: Global methylation analysis, stratified by genomic feature, in placenta.** Boxplots depict distribution of DNAm each diagnostic group of samples, stratified by sex. Number of clusters included in each genomic feature are also shown via 'n' (see Methods). P-values for t-tests comparing different groups. **Panel A**) CpG islands **Panel B**) CpG island shores **Panel C**) CpG island shelves **Panel D**) Open seas

# CHAPTER 5: PLACENTA DNA METHYLATION IS ASSOCIATED WITH FETAL SEX AT *ZNF300*

Shan V. Andrews[1,2], Christine Ladd-Acosta[1,2,3], Kelly M. Bakulski[4], Jason I. Feinberg[2,5], Rakel Tryggvadottir[3], Lisa A. Croen[6], Irva Hertz-Picciotto[7,8], Craig J. Newschaffer[9,10], Ruofan Yao[11], Carolyn M. Salafia[12], Andrew P. Feinberg[3,13], Kasper D. Hansen[3,14,15], M. Daniele Fallin[2,3,5]

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA

[2]Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA

[3]Center for Epigenetics, Institute for Basic Biomedical Sciences, Johns Hopkins School of Medicine, 733 N. Broadway, Baltimore, MD 21205

[4]Department of Epidemiology, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109.

[5]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Baltimore, MD, 21205.

[6]Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612.

[7]Department of Public Health Sciences, School of Medicine, University of California Davis, 4610 X St, Sacramento, CA 95817.

[8]MIND Institute, University of California Davis, 2825 50th St, Sacramento, CA 95817.

[9]AJ Drexel Autism Institute, Drexel University, 3020 Market St #560, Philadelphia, PA 19104.

[10]Department of Epidemiology and Biostatistics, Drexel University Dornsife School of Public Health, 3125 Market St, Philadelphia, PA 19104.

[11]Department of Obstetrics, Gynecology and Reproductive Medicine, University of Maryland School of Medicine, 22 S Greene Street, Baltimore, MD 21201, USA.

[12]Placental Analytics LLC, 187 Overlook Circle, New Rochelle, NY 10984, USA

[13]Department of Medicine, Johns Hopkins School of Medicine, 733 N. Broadway, Baltimore, MD 21205

[14]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA

[15]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, 1800 Orleans Street, Baltimore, MD 21287, USA

**5.1 Abstract**

Background: The placenta undergoes vast epigenomic reprogramming. In normal development, there are sex-specific differences in placenta function and responses to nutritional, stress-related, or environmental insults. While a few gene expression studies have interrogated the mechanism of these sex differences, previous studies of placenta DNA methylation have been limited by either limited methylome coverage or small sample size. To address these limitations, we use a comprehensive measurement of methylation (whole-genome bisulfite sequencing) on 37 ($N_{males}$ = 17, $N_{females}$ = 20) fetal side placenta samples and search for differentially methylated regions (DMRs) according to fetal sex.

Results: We identified one genome wide significant ($p_{permutation}$ = 0.015) DMR where males exhibited on average 15% higher methylation levels in a CpG island promoter of *ZNF300*, a gene previously associated with cell proliferation and tumorogenesis. This fetal sex DMR is placenta-specific, and is consistently replicated with similar magnitude in 7 additional sets of samples (made up of 6 independent studies), including full term and preterm placenta datasets and mixed cell and single cell type datasets, and persists in different disease states. Females can be dichotomized into those who exhibit *ZNF300* methylation levels similar to males and those that

have lower methylation levels. Methylation levels in this DMR are associated with placenta perimeter ($P = 0.044$), area ($P = 0.009$) and maximum diameters ($P = 0.059$); these placenta morphology findings appear to be predominantly among females. Finally, there is suggestive evidence that *ZNF300* DMR methylation levels mediate chromosome 5 and chromosome X SNP influences on these morphological features.

Conclusions: Placenta methylation levels at the *ZNF300* promoter are differential by sex and associated with placenta morphological features. Both of these findings are driven by a subset of female samples that have lower *ZNF300* methylation levels than males.

Keywords: placenta, DNA methylation, fetal sex, whole genome bisulfite sequencing

**5.2 Introduction**

It is well established that there are sex biases in phenotypes across the life course. For example, males are 20% more likely to have a poorer outcome in pregnancies complicated by preeclampsia, preterm delivery, and intrauterine growth restriction (IUGR) (Vatten & Skjaerven, 2004). And later in life, males are more likely to be diagnosed with autism spectrum disorder (ASD) (Christensen et al., 2016), schizophrenia (Aleman, Kahn, & Selten, 2003), and cardiovascular outcomes (Bots, Peters, & Woodward, 2017), and females are more likely to be diagnosed with autoimmune disorders (Ngo, Steyn, & McCombe, 2014). Moreover, it is increasingly appreciated that the developmental and/or gestational period is critical to the programming of disease late in life – the "developmental origins of health and disease" construct (Wadhwa, Buss, Entringer, & Swanson, 2009). The placenta is a key organ during fetal development, providing a critical interface between the mother and fetus, and regulating the transfer of nutrients, waste, and oxygen (Gude, Roberts, Kalionis, & King, 2004). Taken

together, these notions suggest that there are sex-specific placental functions that can ultimately contribute to skewed sex ratios for adverse health outcomes.

Male-female differences have been demonstrated during the gestational period, particularly for adaptations to adverse insults or environments (Clifton, 2010). For example, female growth has been shown to be reduced in the presence of mild maternal asthma; growth can be returned to normal levels with use of inhaled steroids (Murphy et al., 2003). Alternatively males continue normal growth in the presence of maternal asthma and only exhibit signs of affected growth when the asthma reaches acute levels (Murphy, Gibson, Talbot, & Clifton, 2005). Similarly, female fetuses in pregnancies complicated by preeclampsia have shown to exhibit reduced growth, while males tend to grow normally (Stark, Clifton, & Wright, 2009). Such findings have led to hypotheses regarding the differential nature of responses to perturbations during pregnancy by sex: that males aim to maintain growth in the presence of an initial adverse insult, leaving less reserve capacity to handle an additional adverse event. In contrast, females reduce growth in the presence of a poor intrauterine environment, allowing for increased ability to address further adverse events (Clifton, 2010). It is also postulated that females tend to be more at risk for adverse outcomes when perturbations occur in the peri-conception or early gestational periods, affecting trophoblast differentiation. In contrast, male fetuses are more vulnerable in mid-to late gestation, when exposures will affect expansion and function of the definitive placenta as well as nutrient transport (Kalisch-Smith, Simmons, Dickinson, & Moritz, 2017). Finally, placenta immune function displays evidence for sex-specific effects as well; for example mouse studies have shown maternal obesity to be linked to placenta inflammation in late gestation, with males

displaying greater inflammation and macrophage activation than females (Kim, Young, Grattan, & Jasoni, 2014; Tarrade, Panchenko, Junien, & Gabory, 2015).

Beyond these observations, little is understood about the mechanism of sex differences in the placenta. Some gene expression studies have provided suggestions; for example a study investigating sorted placenta cell types found pathways such as graft vs. host disease, immune function, and inflammation to be enriched in genes differentially expressed by sex (Cvitic et al., 2013). Another meta-analysis of placenta microarray data discovered pathways such as cell growth, proliferation, and hormonal function (Buckberry, Bianco-Miotto, Bent, Dekker, & Roberts, 2014). Additional insights into mechanisms of sex-specific function may also come from investigating epigenetic data, or specifically DNA methylation (DNAm). Sex-related mechanisms may be particularly illuminated through the placenta methylome given previous studies highlighting its role in various phenotypes such as preeclampsia (Kulkarni, Chavan-Gautam, Mehendale, Yadav, & Joshi, 2011; Yuen, Peñaherrera, von Dadelszen, McFadden, & Robinson, 2010) and IUGR (Roifman et al., 2016), its sensitivity to environmental exposures during the prenatal period (Maccani et al., 2015; Maccani, Koestler, Houseman, Marsit, & Kelsey, 2013; Mohanty et al., 2015; Suter et al., 2011), and its dynamic nature across gestation (Novakovic et al., 2011). Moreover, there have been many recent studies successfully demonstrating autosomal DNAm differences across a variety of tissue types, such as peripheral blood (Inoshita et al., 2015; Price et al., 2013; Singmann et al., 2015; Suderman et al., 2017), cord blood (Yousefi et al., 2015), pancreas (Hall et al., 2014), prefrontal cortex (Xu et al., 2014), and two recent studies in placenta (Gong et al., 2018; Martin et al., 2017). This first of these previous placenta studies successfully identified 91 significant autosomal sites related to fetal sex

in a small for gestational age cohort, implicating similar pathways to those found by gene expression studies such as immune function, growth/transcription factor signaling, and transport across cell membranes. However, that study measured DNAm via the Illumina 450k microarray (Martin et al., 2017). While this platform is the standard measurement tool for DNAm studies, it is still limited in its coverage of the DNA methylome. The second placenta study addressed these concerns by measuring placenta DNAm via the gold standard technique of whole genome bisulfite sequencing (WGBS), but only studied 4 samples (2 female, 2 male). Beyond sex chromosomes, they observed one autosomal differentially methylated region in the gene body of *CSMD1,* and this was not observed in 8 other tissues, suggesting that it is placenta-specific (Gong et al., 2018).

In this manuscript, we measure placenta DNAm via WGBS on a much larger set of samples ($N_{males}$ = 17, $N_{females}$ = 20) to identify autosomal differential methylation corresponding to fetal sex. We identify a genome-wide significant differentially methylated region (DMR) in the promoter of *ZNF300*. We demonstrate the persistence of this hit in numerous publically available datasets, describe potential evidence for local and X chromosome genetic variants controlling DNAm levels in this region, and describe how both of these molecular features are associated with placenta morphology.

## 5.3 Methods

### *5.3.1 EARLI study description*

The Early Autism Risk Longitudinal Investigation (EARLI) is a prospective, enriched-risk birth cohort that has been described in detail elsewhere (Newschaffer et al., 2012). The EARLI study was approved by Human Subjects Institutional Review Boards (IRBs) from each of the four

study sites (Johns Hopkins University, Drexel University, University of California Davis, and Kaiser Permanente). Informed consent was obtained from all participating families. The 232 mothers with a subsequent child born through this study had births between November 2009 and March 2012. Placental biopsy samples were collected after delivery at each clinical lab site using Baby Tischler Punch Biopsy Forceps. Samples were stored at ambient temperature in RNAlater vials (Qiagen, Cat. No. 76154) and shipped same-day to the Johns Hopkins Biological Repository (JHBR) in Baltimore, Maryland, for storage at -190°C until processing. After biopsy, the placentas were then placed flat in a bag with 10% formalin at a volume roughly equivalent to the size of the placenta. These bags were then shipped to Placenta Analytics, LLC, where they underwent extensive morphological and surface imaging testing.

Infants were followed with extensive neurophenotyping until age three, when a diagnosis of ASD could reliably be made. Children were classified as having autism spectrum disorder (ASD) if they met or exceeded the ASD cutoff of the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) and met the Diagnostic and Statistical Manual of Mental Disorders, 4$^{th}$ edition, Text Revision (DSM-IV-TR) ("DSM-IV: Diagnostic and Statistical Manual of Mental Disorders | JAMA | The JAMA Network," n.d.) criteria for Autistic Disorder or Pervasive Developmental Disorder – Not otherwise specified (PDD-NOS). Children were classified as having non-typical development (NTD) if they did not meet the criteria for ASD classification but had two or more Mullen Scales of Early Learning (Mullen, 1995) subtests ≥ 1.5 SD below the mean and/or one or more Mullen subtests ≥ 2 SD below the mean and/or ADOS ≤ 3 points below the ASD cutoff. Children were classified as having typical development (TD) if they did not meet the criteria for ASD classification and had no more than one Mullen subtest ≥ 1.5 SD

below the mean and no Mullen subtest ≥ 2 SD below the mean and ADOS > 3 points below the ASD cutoff.

*5.3.2 Whole genome bisulfite sequencing processing, alignment, and quality control*

Genomic DNA (gDNA) from biopsy punches was extracted at JHBR using a QIAgen QIAsymphony automated workstation with the Blood 1000 protocol of the DSP DNA Midi kit (Cat. No. 937255) as per manufacturer's instructions. 250ng gDNA was bisulfite treated using the Zymo EZ-96 DNA Methylation Kit (Cat. No. D5004), and sequencing libraries were made with the NEBNext Ultra DNA Library Prep Kit for Illumina. Library preparation was performed at the Center for Epigenetics and libraries were sequenced at ~4x coverage at the High Throughput Sequencing Center at the Johns Hopkins University School of Medicine using the 3 Illumina HiSeq 2500 instruments. Initially 133 placenta samples were run on 16 flow cells; two samples failed due to a fluidics error and were re-run using a 17$^{th}$ flow cell.

Raw .fastq files were aligned the human reference genome (hg38) downloaded from Ensembl (Aken et al., 2017) via Bismark (v0.16.3) (Krueger & Andrews, 2011) using the default arguments for paired-end sequencing data. M-bias plots (Hansen, Langmead, & Irizarry, 2012) were used to inform the trimming of the 5bp of both the first and second reads, and methylation status was called using the Bismark methylation extractor function. All subsequent analyses used R (version 3.4.2) and the R package *bsseq* (version 1.13.6) ("bsseq," n.d.). The bsseq function *read.bismark()* was used to read in Bismark output files ("cytosine report" format) and obtain DNAm measurements on a per-CpG site basis. A total of 29,091,077 CpG sites were measured with at least 1 read in 1 sample out of the total 133. Next methylation levels were smoothed using the BSmooth algorithm that has been previously described (Hansen et al., 2012). Briefly,

this technique employs a local-likelihood smoothing algorithm to recapitulate higher coverage whole genome bisulfite sequencing data from lower coverage data, exploiting known local correlation structure in DNAm data. We used the *BSmooth()* function in the *bsseq* R package with the default arguments.

### *5.3.3 Identification of differentially methylated regions*

To characterize normal placentas, we limited our initial discovery sample to 37 samples ($N_{males}$ = 17, $N_{females}$ = 20) with a classification of TD at 3 years of age. We removed CpGs with little or no coverage in a large proportion of samples to avoid detection of DMRs in these regions, as they are likely to be false positives ("bsseq," n.d.). Specifically we limited DNAm data to sites at which at least 8 males and 10 females had ≥ 2 reads at each CpG site, and also removed sex chromosome CpGs, leaving a total of 26,157,032 CpG sites. Finally we used the *BSmooth.tstat.fix()* and *dmrFinder()* functions in *bsseq* with the default arguments to identify DMRs. Statistical significance was assessed by creating 1000 null sets of methylation data via scrambling sample labels for fetal sex and re-running the DMR-finding algorithm; p-values were calculated as the number of null sets in which at least 1 DMR had a ≥ 'areaStat' value (sum of t-statistics) and width (in bp) than the candidate DMR.

### *5.3.4 Placenta replication data sets*

After identifying a genome-wide significant DMR near *ZNF300*, we sought to examine the extent to which this difference replicated in additional datasets. First, we examined the DMR in the 50 EARLI WGBS samples classified as NTD ($N_{males}$ = 26, $N_{females}$ = 24), termed Placenta Replication 1 (PR 1). Next we downloaded the manifest for the Illumina HumanMethylation450 BeadChip (450K), the primary platform for which publically available placenta DNAm data are

available, and found that 10 450K probes overlapped with the *ZNF300* DMR coordinates. Therefore, we first downloaded 3 full term, mixed cell placenta datasets (matching the type of data we had for EARLI) from the Gene Expression Omnibus (GEO) (Edgar, Domrachev, & Lash, 2002) (GSE75248, GSE71678, and GSE75196), and obtained an additional dataset of this kind (Roifman et al., 2016) from the study authors; these were termed PRs 2-5 respectively. Next to investigate if this DMR was present earlier in the gestational period, we downloaded the preterm samples from an additional dataset (GSE57767), which we termed PR 6. Finally, to preclude the possibility that this DMR was driven by cell type heterogeneity (Jaffe & Irizarry, 2014), we downloaded an additional 450k dataset of a single placenta cell type, villous cytotrophoblasts (GSE593208), which we termed PR 7.

For 450k datasets for which raw data were available (PRs 2-5 and 7), we implemented the following processing and QC steps: functional normalization (Fortin et al., 2014), removal of samples with low overall intensity (median methylated or unmethylated signal < 11) or with a detection p-value > 0.01 in more than 1% of probes, and removal of probes described as ambiguously mapping (Chen et al., 2013), or with a detection p-value > 0.01 in more than 10% of samples. For PR 6, for which raw data were not available, we performed quantile normalization (Touleimat & Tost, 2012) and removed ambiguously mapping probes.

*5.3.5 Additional tissue replication data sets*

We also sought to discover the extent to which the genome-wide significant placenta DMR was present in different tissue types; we examined 450k data from cord blood, fetal brain, and peripheral blood. The cord blood data were obtained from EARLI samples; processing and quality control has been previously described (Andrews et al., 2017). Normalized percent

methylation values for fetal brain data were downloaded from GEO (GSE58885). Peripheral blood data from 3-5 year old children were used from the Study to Explore Early Development; processing and quality control has also been previously described (Andrews et al., 2017).

*5.3.6 Genotype processing and methylation quantitative trait loci query*

We were first interested in quantifying the relationship of chromosome 5 SNPs to the *ZNF300* DMR methylation levels. Genotype processing and QC for EARLI autosomal SNPs has been previously described (Andrews et al., 2017). Imputed genotype data were available on 682,195 SNPS for 121 of the 133 samples on which WGBS data existed. We imposed a 5% MAF threshold and removed duplicate positions, which limited the number of SNPs to 451,693, and then lifted the data from hg19 to hg38 (to match genome build on which WGBS alignment was performed) using the liftOver software (Hinrichs et al., 2006), yielding a final count of 451,680 SNPs on chromosome 5. We were also interest in quantifying X chromosome SNP relationships to the DMR methylation levels. We measured 108,563 SNPs using the Illumina Omni5 plus exome array, which included 105,121 SNPs from the non-pseudoautosomal (nonPAR) region and 3,442 SNPs from the pseudoautosomal (PAR) region. Imposing a 5% MAF threshold yielded a total of 47,399 SNPs (44,992 nonPAR and 2,407 PAR), and lifting genome builds left a final total of 47,265 SNPs.

To perform the methylation quantitative trait loci (meQTL) queries, we regressed the mean methylation level in the *ZNF300* DMR onto each genotype (using an additive model) adjusting for the first four principal components of ancestry and gestational age. For the X chromosome SNPs, we conducted this analysis in each sex separately ($N_{males} = 70$, $N_{females} = 50$) and then

meta-analyzed those results together using the METAL software (Willer, Li, & Abecasis, 2010) under the default conditions.

*5.3.7 Morphology mediation analyses*

We regressed mean *ZNF300* methylation levels onto each morphological feature, adjusting for gestational age and birthweight; we did this analysis first in all samples on which DNAm and morphology data were available (N = 56), and then repeated the analyses in males (N = 34) and females (N = 22) separately. Finally, to explore the possibility that the suggestive genotype effects of meQTLs influenced these morphology phenotypes through *ZNF300* DNAm, we employed a causal inference testing (CIT), or mediation analysis, approach as we previously implemented for rheumatoid arthritis and child allergy (Hong et al., 2015; Liu et al., 2013). Briefly, this method conducts a series of regressions to establish the genotypic effect on a phenotype that acts through methylation, exploiting the implicit nature of the SNP effect to precede all other effects temporally. These regressions are:

1) Regression of phenotype (Y) onto methylation (M): Y and M are associated but no causality established
2) Regression of methylation (M) onto genotype (G): G is causally associated with M
3) Regression of Y onto G: G is causally associated with Y
4) Regression of Y onto G, adjusting for M: Attenuation of beta coefficient indicates that the causal effect of G on Y is mediated by M

We conducted a series of CIT procedures using all SNPs from the chromosome 5 and chromosome X meQTL queries that passed a suggestive significance threshold of $p \leq 1E\text{-}4$ (n = 17), the mean methylation level from the *ZNF300* DMR, and the 4 placenta morphology

phenotypes. For each regression, we included the total number of samples which had measurements for the form of data included in that regression. For regression 1, this was 56 samples. For regression 2, this was 121 samples. For regressions 3 and 4, this was 51 samples. For all regressions involving X chromosome genotype data, we conducted analyses separately in males and females and then meta-analyzed the results as done for the meQTL query.

## 5.4 Results

### *5.4.1 Whole genome bisulfite sequencing study of fetal sex and replication*

A single DMR met a genome-wide permutation-base p-value threshold < 0.05: a 1597 bp region on chromosome 5 ($p_{permutation}$ = 0.015; **Supplementary Figure V-1**). This DMR is characterized by 15% hypermethylation on average in male compared to female placentas (**Figure 5.1A**), and is not driven by fetal sex-related coverage of the region (**Figure 5.1B**). It encompasses a roughly 1.5 Kb region in a CpG island promoter of the *ZNF300*, and overlaps placenta-specific DNaseI hypersensitive sites identified in both male and female samples (**Figure 5.1C**). The DMR also overlaps 56 transcription factor binding sites (TFBSs), 53 of which map to the promoter region of the *ZNF300*. However, this number of TFBSs does not greatly distinguish *ZNF300* from other genes in the genome, as roughly 30% of genes exhibit higher numbers of promoter-targeting TFBSs (**Supplementary Figure V-2**).

We next sought to examine the extent to which the *ZNF300* fetal sex DMR replicated in additional datasets. First we examined EARLI placentas from independent children who went on to be classified as non-typically developing ($N_{males}$ = 26, $N_{females}$ = 24); these samples also showed a sex DMR at ZNF300, with a mean male-female difference (MD) of 0.12 (Placenta Replication or PR 1 – **Figure 5.2A**). We also examined Illumina 450K Array data from 6

additional datasets, using multiple probes on the array that overlapped the DMR region. Four datasets from full term, mixed cell samples replicated the sex DMR, with MD values between 0.12 and 0.22 (PR 2-5 – **Figure 5.2B-E**). A separate dataset of preterm, mixed cell samples showed a MD value of 0.16 (PR 6 – **Figure 5.2F**). Finally, a sample of preterm placentas using a single placenta cell type (villous cytotrophoblasts) replicated with a MD value of 0.25 (PR 7 – **Figure 5.2G**). In addition to demonstrating the robustness of this DMR to gestational period (PR 6 & 7) and potential confounding by cell type heterogeneity (PR 7), these replication sets also demonstrate the persistence of this DMR in different disease states. This persistence is true of early life phenotypes such as preeclampsia (cases and controls – PR 4; cases only – PR 6) and intrauterine growth restriction (PR 5), as well as later in life phenotypes such as non-typical neurodevelopment (PR 1).

*5.4.2 Non-placenta tissues*

Differential methylation by sex was not observed at *ZNF300* in cord blood samples, fetal brain or peripheral blood samples. Mean differences between males and females were 0.004, 0.016, and 0.01, respectively (**Figure 5.3**).

*5.4.3 Growth relationship with placenta DNAm at ZNF300*

This promoter region of *ZNF300* had previously been identified for differential placenta DNAm related IUGR (Roifman et al., 2016). Specifically, in monochorionic IUGR discordant twin samples, placentas from IUGR samples were hypermethylated in this region. In our EARLI samples, and one of the publically available replication samples (PR 3), we derived a small-for-gestational age (SGA) phenotype as a proxy for IUGR, based on birthweight and gestational age. In EARLI samples, hypermethylation in SGA samples was not seen across male and female

samples but was seen when looking at females only. In males, the opposite effect of hypomethylation was seen for the SGA samples (**Supplementary Figure V-3A-C**). In PR3, hypomethylation of SGA samples was seen when looking across all samples and in males only, whereas no relationship was seen in females only (**Supplementary Figure V-3D-F**).

*5.4.4 The association between ZNF300 methylation levels and placenta morphological features*

We sought to understand the potential relationship between methylation levels in the *ZNF300* DMR and placenta morphological features, specifically placenta weight and placenta perimeter, area, and max diameter as determined by 2D fetal surface imaging. Across male and female samples on which methylation and morphology data were both available (N = 56), average methylation levels in the DMR were significantly associated with placenta perimeter ($p = 0.044$), area ($p = 0.009$), and max diameter ($p = 0.059$), but not weight ($p = 0.714$), adjusted for birth weight and gestational age (**Figure 5.4**). Sex-stratified analyses revealed that this association was driven by female samples (**Supplementary Figure V-4**). The *ZNF300* methylation association to placenta morphology was not seen in males.

*5.4.5 Integrating genotype, ZNF300 methylation levels, and placenta morphology*

Finally, we sought to integrate genotype, methylation, and morphology information to investigate the potential role of *ZNF300* methylation levels as a mediator of *cis* (chromosome 5) or *trans* (chromosome X) SNP effects on morphology. 17 SNPs were suggestively associated ($p \leq 1E-4$) with *ZNF300* methylation levels (6 from chromosome 5 and 11 from chromosome X; **Supplementary Figure V-5**). Of these, one SNP on chromosome 5 and 7 SNPs on chromosome X (all harbored in *GK*) displayed evidence for a causal effect on placenta perimeter, area, and

max diameter that was partially mediated through *ZNF300* methylation levels (**Figure 5** and **Supplementary Figures V-6** to **V-9**).

### 5.5 Discussion

This study uses WGBS-measured DNAm data from fetal side placenta samples to identify a DMR associated with fetal sex harbored in a CpG island promoter region of *ZNF300*. Additional datasets demonstrate the replication of this DMR across multiple studies at a very similar magnitude (~15% higher methylation in males), and provide evidence for the persistence of the DMR at different gestational periods, in a single placenta cell type, and in different disease states. The DMR is placenta-specific, cannot be attributed to potential confounding variables such as gestational age, mode of delivery, or race, and does not share a significant degree of homology with sex chromosome sequences. Suggestive evidence exists for genetic drivers of *ZNF300* methylation levels, which in turn are related to placenta morphology.

Interestingly, the *ZNF300* sex DMR is characterized by greater variability in female samples than males, as female samples cluster into a group with DNAm levels similar to those of males and another group with lower methylation levels who drive differential mean DNAm levels. Lower DNAm levels in the identified region are associated with smaller placenta perimeters, areas, and maximum diameters, but not placenta weights. This finding is driven by a female samples, and potentially reflects mediation of SNP influences on placenta morphology by DNAm.

The chief strength of our finding is the extensive replication of the DMR across 7 different data sets, including 6 independent studies. To our knowledge, this is the strongest body of evidence for an autosomal sex-related DNAm difference that has been demonstrated to date.

In addition, two lines of evidence indicate that the DMR is placenta-specific. First, direct lookup of the identified *ZNF300* region in DNAm isolated from cord blood, peripheral blood, and fetal tissues showed no relationship to sex. Second, the vast majority of previous studies of autosomal DNAm differences related to sex in other tissues, such as across peripheral blood (Inoshita et al., 2015; Price et al., 2013; Suderman et al., 2017), cord blood (Yousefi et al., 2015), pancreas (Hall et al., 2014), and brain (Xu et al., 2014) did not report *ZNF300* differences. One study in peripheral blood did report 5 Bonferroni significant *ZNF300* sites in their initial discovery phase, but only replicated the result in 1 of their 3 replication cohorts, and only for 2 sites (Singmann et al., 2015).

*ZNF300* functions as a transcriptional repressor, and hypermethylation in the promoter of this gene has been previously associated with decreased gene expression (Zhao et al., 2014). Expression of the *ZNF300* has been previously associated with tumorogenesis in HeLa cells (Wang et al., 2012), but knockdown of *ZNF300* has been previously associated with cell proliferation in K562 cells (Cai et al., 2014). Our results indicated that placenta cells (chiefly whichever cell types drive this methylation association) tend to mimic K562 cells in their *ZNF300* functionality, as hypomethylated female samples (thus those with increased *ZNF300* expression) tended to have placentas characterized by smaller areas, perimeters, and max diameters, indicating a decreased proliferative capacity.

A previous study of placenta DNAm also identified the *ZNF300* promoter region as being differentially methylation in monochorionic twin samples discordant for IUGR (Roifman et al., 2016). Using SGA as a surrogate for IUGR, we did not replicate this finding using EARLI samples or a publically available dataset. However, our findings were made from a limited

number of SGA samples and should be investigated further, particularly in light of our findings relating DNAm in this region to placenta morphology. Of note is that while we observed hypermethylation at *ZNF300* in SGA samples (consistent in direction with IUGR as reported in the previous paper), when limiting analyses to females only, we also found lower DNAm levels in this region to be associated with smaller placenta areas, even though it has been previously reported that small placenta size is associated with IUGR in women with low pregnancy-associated plasma protein-A (Proctor et al., 2009).

The hypomethylated female samples represented one mode of female samples in the identified ZNF300 region. While this group drove the differential methylation and morphology findings, the other group of females had methylation levels that tended to cluster with male samples. In other words, females were characterized by a greater degree of DNAm variability in this region compared to males, and this region may have well been discovered by screens for sex-specific differential DNAm variability. The notion of greater DNAm variability in female samples, especially in this case where there are downstream morphological consequences, is particularly intriguing in the context of the placenta. It has been theorized that faster male growth during gestation comes at the risk of decreased plasticity or robustness to perturbations or adverse conditions during pregnancy (Clifton, 2010; Eriksson, Kajantie, Osmond, Thornburg, & Barker, 2010). The greater distribution of placenta *ZNF300* methylation levels and their linked morphological phenotypes that was observed in females in our study is congruent with this notion of female adaptability. Furthermore, there was a strong association between *ZNF300* methylation levels and these phenotypes when limiting analyses to females only, while no relationship was seen in males only. Males instead had little variability in *ZNF300* DNAm levels.

Future work should continue to investigate sex-related mean differences in DNAm as we did here in light of our small discovery sample size (N = 37). However, given the nature of the *ZNF300* DMR discovered in our study, sex-focused studies of placenta DNAm should screen for differential variability as well. Future work specific to the ZNF300 DMR should first aim to better characterize the source(s) of this methylation difference. One potential area of investigation could be the role of sex hormones; this path is particularly intriguing given previous findings linking peripheral blood-derived DNAm sex differences in early life and sex hormone levels (Suderman et al., 2017). We also demonstrated suggestive evidence for local and more distant genetic effects on ZNF300 DNAm levels; these findings should be confirmed in greater sample sizes of joint genotype and DNAm data than were available herein (N = 121).

Our evidence demonstrating the association between *ZNF300* DNAm levels and placenta morphological phenotypes, including the suggestive genotype relationships acting through methylation to influence placenta morphology, should also be confirmed in greater sample sizes. This work should include efforts to obtain more direct evidence for the presence of this DMR during the gestational period, to better argue for its functional relevance. Finally, specific investigations to characterize the biomodality of female *ZNF300* DNAm levels should be conducted to understand the factors driving females to cluster in one group vs. another, to what extent this grouping characterizes DNAm levels across the genome, and to discover if these two groups differ in early and/or later in life health outcomes.

### 5.6 Conclusions

We demonstrate a sex-related DMR identified in placenta samples that is harbored in a CpG island promoter of *ZNF300*. This placenta-specific DMR is consistently replicated across

different studies with a similar magnitude, and is robust to gestational period, cell type heterogeneity, and disease state. Female samples with lower DNAm levels in this region have smaller placenta areas, perimeters, and maximum diameters than females with DNAm levels similar to males. These DNAm influences on placenta morphology may be mediating local and more distant (X chromosome) genetic effects. This DMR provides insight into potential molecular mechanisms of known sex differences in placenta structure and function. We argue for the further characterization of this DMR for its role in contributing to sex-biased phenotypes later in life.

### 5.7 References


Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., … Flicek, P. (2017). Ensembl 2017. *Nucleic Acids Research*, *45*(D1), D635–D642. https://doi.org/10.1093/nar/gkw1104

Aleman, A., Kahn, R. S., & Selten, J.-P. (2003). Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Archives of General Psychiatry*, *60*(6), 565–571. https://doi.org/10.1001/archpsyc.60.6.565

Andrews, S. V., Ellis, S. E., Bakulski, K. M., Sheppard, B., Croen, L. A., Hertz-Picciotto, I., … Fallin, M. D. (2017). Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. *Nature Communications*, *8*(1), 1011. https://doi.org/10.1038/s41467-017-00868-y

Bots, S. H., Peters, S. A. E., & Woodward, M. (2017). Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health*, *2*(2), e000298. https://doi.org/10.1136/bmjgh-2017-000298

bsseq. (n.d.). Retrieved February 15, 2018, from http://bioconductor.org/packages/bsseq/

Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A., & Roberts, C. T. (2014). Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. *Molecular Human Reproduction*, *20*(8), 810–819. https://doi.org/10.1093/molehr/gau035

Cai, J., Gong, R., Yan, F., Yu, C., Liu, L., Wang, W., … Huang, Z. (2014). ZNF300 knockdown inhibits forced megakaryocytic differentiation by phorbol and erythrocytic differentiation by

arabinofuranosyl cytidine in K562 cells. *PloS One*, *9*(12), e114768. https://doi.org/10.1371/journal.pone.0114768

Chen, Y., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., … Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, *8*(2), 203–209. https://doi.org/10.4161/epi.23470

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Centers for Disease Control and Prevention (CDC). (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years--Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Clifton, V. L. (2010). Review: Sex and the human placenta: mediating differential strategies of fetal growth and survival. *Placenta*, *31 Suppl*, S33-39. https://doi.org/10.1016/j.placenta.2009.11.010

Cvitic, S., Longtine, M. S., Hackl, H., Wagner, K., Nelson, M. D., Desoye, G., & Hiden, U. (2013). The human placental sexome differs between trophoblast epithelium and villous vessel endothelium. *PloS One*, *8*(10), e79233. https://doi.org/10.1371/journal.pone.0079233

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders | JAMA | The JAMA Network. (n.d.). Retrieved February 15, 2018, from https://jamanetwork.com/journals/jama/article-abstract/379036?redirect=true

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–210.

Eriksson, J. G., Kajantie, E., Osmond, C., Thornburg, K., & Barker, D. J. P. (2010). Boys live dangerously in the womb. *American Journal of Human Biology: The Official Journal of the Human Biology Council*, *22*(3), 330–335. https://doi.org/10.1002/ajhb.20995

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., … Hansen, K. D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, *15*(12), 503. https://doi.org/10.1186/s13059-014-0503-2

Gong, S., Johnson, M. D., Dopierala, J., Gaccioli, F., Sovio, U., Constância, M., … Charnock-Jones, D. S. (2018). Genome-wide oxidative bisulfite sequencing identifies sex-specific methylation differences in the human placenta. *Epigenetics*, 1–32. https://doi.org/10.1080/15592294.2018.1429857

Gude, N. M., Roberts, C. T., Kalionis, B., & King, R. G. (2004). Growth and function of the normal human placenta. *Thrombosis Research*, *114*(5–6), 397–407. https://doi.org/10.1016/j.thromres.2004.06.038

Hall, E., Volkov, P., Dayeh, T., Esguerra, J. L. S., Salö, S., Eliasson, L., … Ling, C. (2014). Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biology*, *15*(12), 522. https://doi.org/10.1186/s13059-014-0522-z

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, *13*(10), R83. https://doi.org/10.1186/gb-2012-13-10-r83

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., … Kent, W. J. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*, *34*(Database issue), D590-598. https://doi.org/10.1093/nar/gkj144

Hong, X., Hao, K., Ladd-Acosta, C., Hansen, K. D., Tsai, H.-J., Liu, X., … Wang, X. (2015). Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nature Communications*, *6*, 6304. https://doi.org/10.1038/ncomms7304

Inoshita, M., Numata, S., Tajima, A., Kinoshita, M., Umehara, H., Yamamori, H., … Ohmori, T. (2015). Sex differences of leukocytes DNA methylation adjusted for estimated cellular proportions. *Biology of Sex Differences*, *6*, 11. https://doi.org/10.1186/s13293-015-0029-7

Jaffe, A. E., & Irizarry, R. A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology*, *15*(2), R31. https://doi.org/10.1186/gb-2014-15-2-r31

Kalisch-Smith, J. I., Simmons, D. G., Dickinson, H., & Moritz, K. M. (2017). Review: Sexual dimorphism in the formation, function and adaptation of the placenta. *Placenta*, *54*, 10–16. https://doi.org/10.1016/j.placenta.2016.12.008

Kim, D. W., Young, S. L., Grattan, D. R., & Jasoni, C. L. (2014). Obesity during pregnancy disrupts placental morphology, cell proliferation, and inflammation in a sex-specific manner across gestation in the mouse. *Biology of Reproduction*, *90*(6), 130. https://doi.org/10.1095/biolreprod.113.117259

Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Kulkarni, A., Chavan-Gautam, P., Mehendale, S., Yadav, H., & Joshi, S. (2011). Global DNA methylation patterns in placenta and its association with maternal hypertension in pre-eclampsia. *DNA and Cell Biology*, *30*(2), 79–84. https://doi.org/10.1089/dna.2010.1084

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., … Feinberg, A. P. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology*, *31*(2), 142–147. https://doi.org/10.1038/nbt.2487

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Maccani, J. Z. J., Koestler, D. C., Houseman, E. A., Marsit, C. J., & Kelsey, K. T. (2013). Placental DNA methylation alterations associated with maternal tobacco smoking at the RUNX3 gene are also associated with gestational age. *Epigenomics*, *5*(6), 619–630. https://doi.org/10.2217/epi.13.63

Maccani, J. Z. J., Koestler, D. C., Lester, B., Houseman, E. A., Armstrong, D. A., Kelsey, K. T., & Marsit, C. J. (2015). Placental DNA Methylation Related to Both Infant Toenail Mercury and Adverse Neurobehavioral Outcomes. *Environmental Health Perspectives*, *123*(7), 723–729. https://doi.org/10.1289/ehp.1408561

Martin, E., Smeester, L., Bommarito, P. A., Grace, M. R., Boggess, K., Kuban, K., … Fry, R. C. (2017). Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period. *Epigenomics*, *9*(3), 267–278. https://doi.org/10.2217/epi-2016-0132

Mohanty, A. F., Farin, F. M., Bammler, T. K., MacDonald, J. W., Afsharinejad, Z., Burbacher, T. M., … Enquobahrie, D. A. (2015). Infant sex-specific placental cadmium and DNA methylation associations. *Environmental Research*, *138*, 74–81. https://doi.org/10.1016/j.envres.2015.02.004

Mullen, E. M. (1995). Mullen Scales of Early Learning. *American Guidance Service Inc.*

Murphy, V. E., Gibson, P. G., Giles, W. B., Zakar, T., Smith, R., Bisits, A. M., … Clifton, V. L. (2003). Maternal asthma is associated with reduced female fetal growth. *American Journal of Respiratory and Critical Care Medicine*, *168*(11), 1317–1323. https://doi.org/10.1164/rccm.200303-374OC

Murphy, V. E., Gibson, P., Talbot, P. I., & Clifton, V. L. (2005). Severe asthma exacerbations during pregnancy. *Obstetrics and Gynecology*, *106*(5 Pt 1), 1046–1054. https://doi.org/10.1097/01.AOG.0000185281.21716.02

Newschaffer, C. J., Croen, L. A., Fallin, M. D., Hertz-Picciotto, I., Nguyen, D. V., Lee, N. L., … Shedd-Wise, K. M. (2012). Infant siblings and the investigation of autism risk factors. *Journal of Neurodevelopmental Disorders*, *4*(1), 7. https://doi.org/10.1186/1866-1955-4-7

Ngo, S. T., Steyn, F. J., & McCombe, P. A. (2014). Gender differences in autoimmune disease. *Frontiers in Neuroendocrinology*, *35*(3), 347–369. https://doi.org/10.1016/j.yfrne.2014.04.004

Novakovic, B., Yuen, R. K., Gordon, L., Penaherrera, M. S., Sharkey, A., Moffett, A., … Saffery, R. (2011). Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. *BMC Genomics*, *12*, 529. https://doi.org/10.1186/1471-2164-12-529

Price, M. E., Cotton, A. M., Lam, L. L., Farré, P., Emberly, E., Brown, C. J., … Kobor, M. S. (2013). Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics & Chromatin*, *6*(1), 4. https://doi.org/10.1186/1756-8935-6-4

Proctor, L. K., Toal, M., Keating, S., Chitayat, D., Okun, N., Windrim, R. C., … Kingdom, J. C. P. (2009). Placental size and the prediction of severe early-onset intrauterine growth restriction in women with low pregnancy-associated plasma protein-A. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, *34*(3), 274–282. https://doi.org/10.1002/uog.7308

Roifman, M., Choufani, S., Turinsky, A. L., Drewlo, S., Keating, S., Brudno, M., … Weksberg, R. (2016). Genome-wide placental DNA methylation analysis of severely growth-discordant monochorionic twins reveals novel epigenetic targets for intrauterine growth restriction. *Clinical Epigenetics*, *8*, 70. https://doi.org/10.1186/s13148-016-0238-x

Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.-Y., … Halperin, E. (2015). Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics & Chromatin*, *8*, 43. https://doi.org/10.1186/s13072-015-0035-3

Stark, M. J., Clifton, V. L., & Wright, I. M. R. (2009). Neonates born to mothers with preeclampsia exhibit sex-specific alterations in microvascular function. *Pediatric Research*, *65*(3), 292–295.

Suderman, M., Simpkin, A., Sharp, G. C., Gaunt, T., Lyttleton, O., McArdle, W., … Relton, C. L. (2017). Sex-associated autosomal DNA methylation differences are wide-spread and stable throughout childhood. *bioRxiv*, 118265. https://doi.org/10.1101/118265

Suter, M., Ma, J., Harris, A., Patterson, L., Brown, K. A., Shope, C., … Aagaard-Tillery, K. M. (2011). Maternal tobacco use modestly alters correlated epigenome-wide placental DNA

methylation and gene expression. *Epigenetics*, *6*(11), 1284–1294. https://doi.org/10.4161/epi.6.11.17819

Tarrade, A., Panchenko, P., Junien, C., & Gabory, A. (2015). Placental contribution to nutritional programming of health and diseases: epigenetics and sexual dimorphism. *The Journal of Experimental Biology*, *218*(Pt 1), 50–58. https://doi.org/10.1242/jeb.110320

Touleimat, N., & Tost, J. (2012). Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*, *4*(3), 325–341. https://doi.org/10.2217/epi.12.21

Vatten, L. J., & Skjaerven, R. (2004). Offspring sex and pregnancy outcome by length of gestation. *Early Human Development*, *76*(1), 47–54.

Wadhwa, P. D., Buss, C., Entringer, S., & Swanson, J. M. (2009). Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Seminars in Reproductive Medicine*, *27*(5), 358–368. https://doi.org/10.1055/s-0029-1237424

Wang, T., Wang, X., Xu, J., Wu, X.-P., Qiu, H., Yi, H., & Li, W.-X. (2012). Overexpression of the human ZNF300 gene enhances growth and metastasis of cancer cells through activating NF-kB pathway. *Journal of Cellular and Molecular Medicine*, *16*(5), 1134–1145. https://doi.org/10.1111/j.1582-4934.2011.01388.x

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Xu, H., Wang, F., Liu, Y., Yu, Y., Gelernter, J., & Zhang, H. (2014). Sex-biased methylome and transcriptome in human prefrontal cortex. *Human Molecular Genetics*, *23*(5), 1260–1270. https://doi.org/10.1093/hmg/ddt516

Yousefi, P., Huen, K., Davé, V., Barcellos, L., Eskenazi, B., & Holland, N. (2015). Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics*, *16*, 911. https://doi.org/10.1186/s12864-015-2034-y

Yuen, R. K., Peñaherrera, M. S., von Dadelszen, P., McFadden, D. E., & Robinson, W. P. (2010). DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. *European Journal of Human Genetics: EJHG*, *18*(9), 1006–1012. https://doi.org/10.1038/ejhg.2010.63

Zhao, Y., Xue, F., Sun, J., Guo, S., Zhang, H., Qiu, B., … Xia, Q. (2014). Genome-wide methylation profiling of the different stages of hepatitis B virus-related hepatocellular carcinoma development in plasma cell-free DNA reveals potential biomarkers for early detection and high-

risk monitoring of hepatocellular carcinoma. *Clinical Epigenetics*, *6*(1), 30. https://doi.org/10.1186/1868-7083-6-30

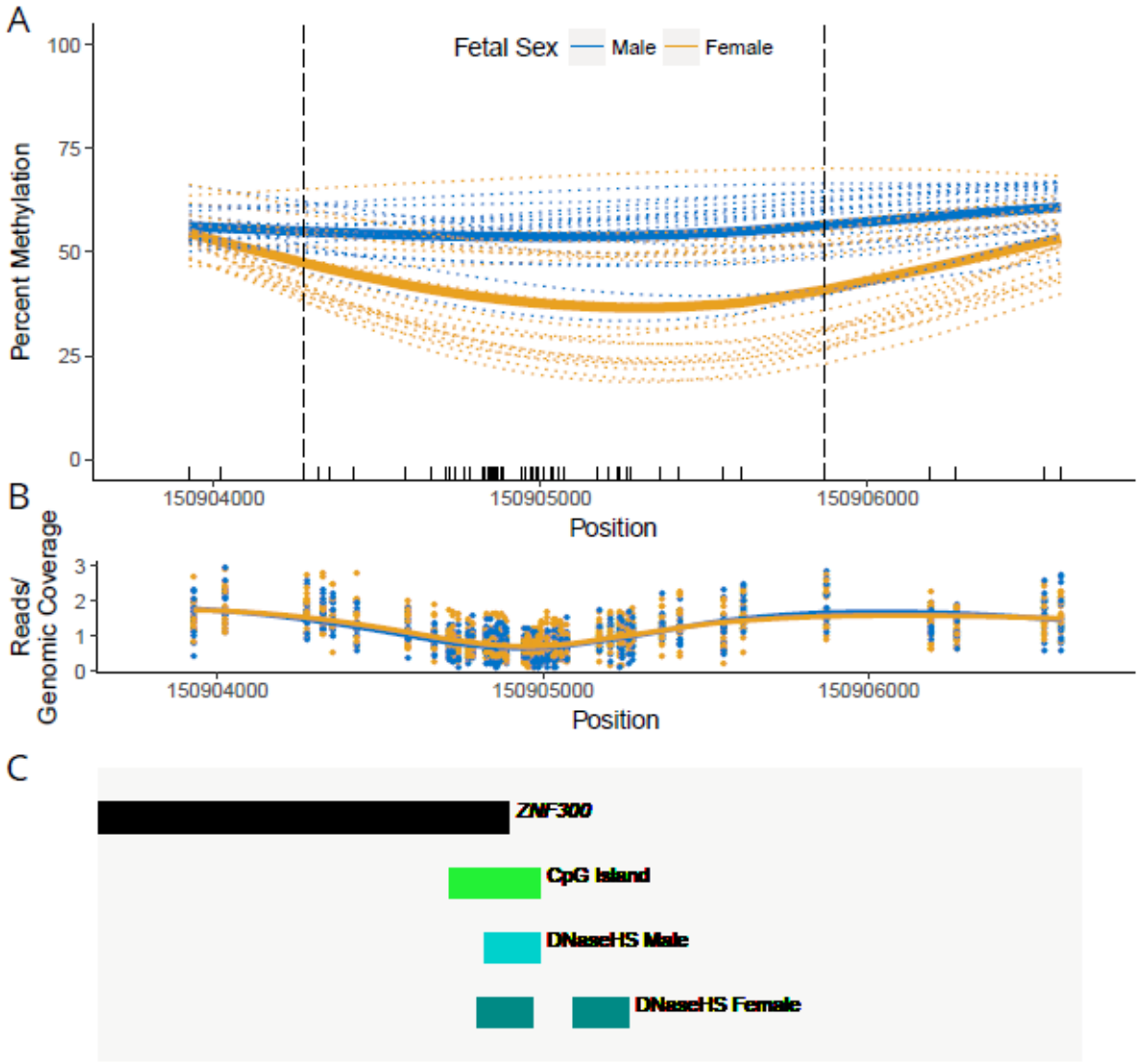



**Figure 5.1: Differential placenta methylation by fetal sex identified at *ZNF300* CpG island promoter. Panel A)** Percent methylation (y-axis) plotted against genomic position (x-axis). Dashed lines represent individual samples and solid lines represent average for male (blue) and female (yellow) samples, respectively. Vertical dashed lines denote actual DMR start and stop coordinates. **Panel B)** Read counts divided by average coverage across all autosomal loci (y-axis) vs. genomic position (x-axis). No differences exist between males and females in the DMR region. **Panel C)** Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

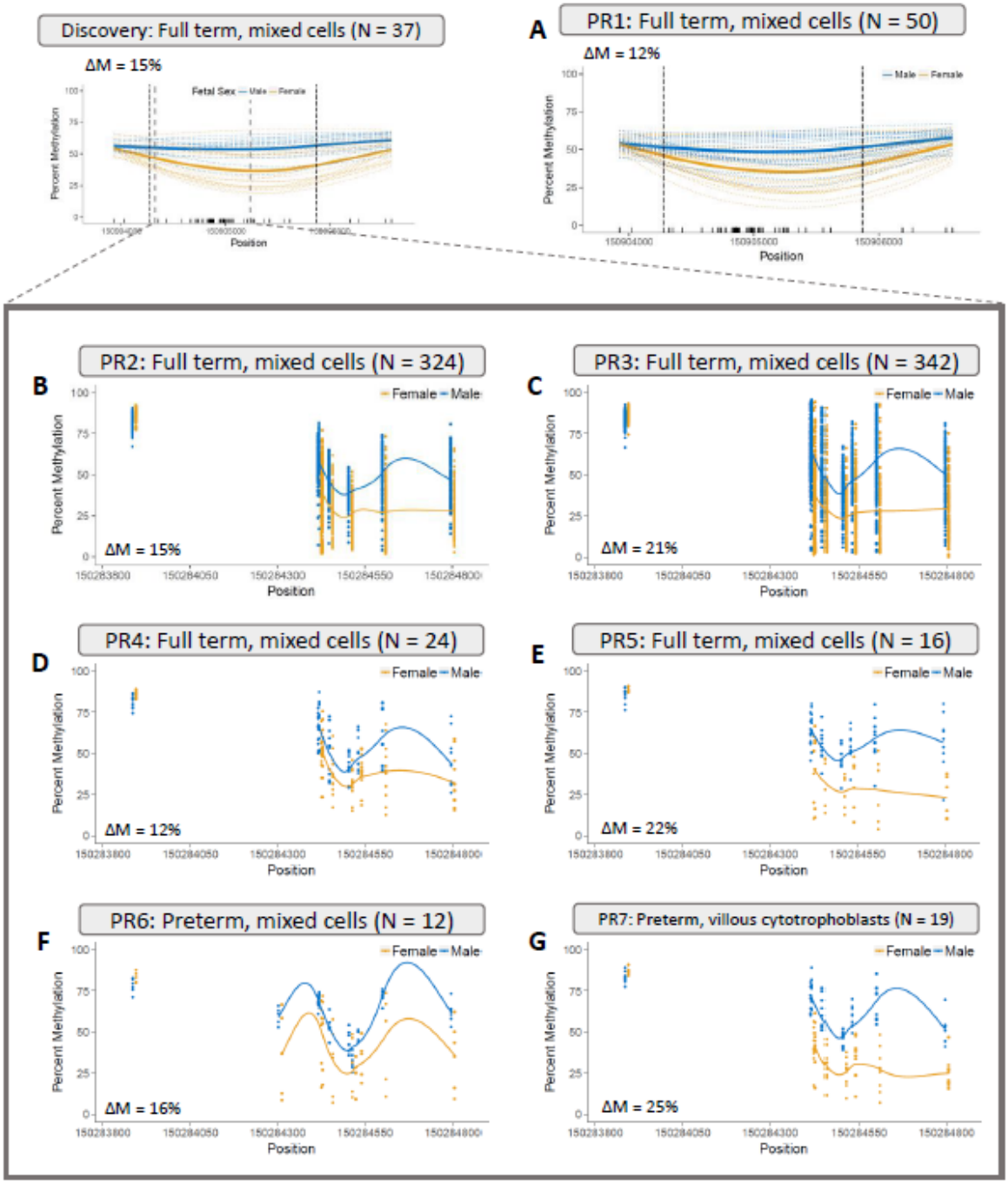



**Figure 5.2: Replication of male hypermethylation at *ZNF300* CpG island promoter. Panel A)** EARLI WGBS non-typically developing samples. **Panels B-G**: 450k replication datasets including preterm (Panels F & G) and single cell type (Panel G) datasets. 450k coverage encompasses a portion of the WGBS-identified DMR.

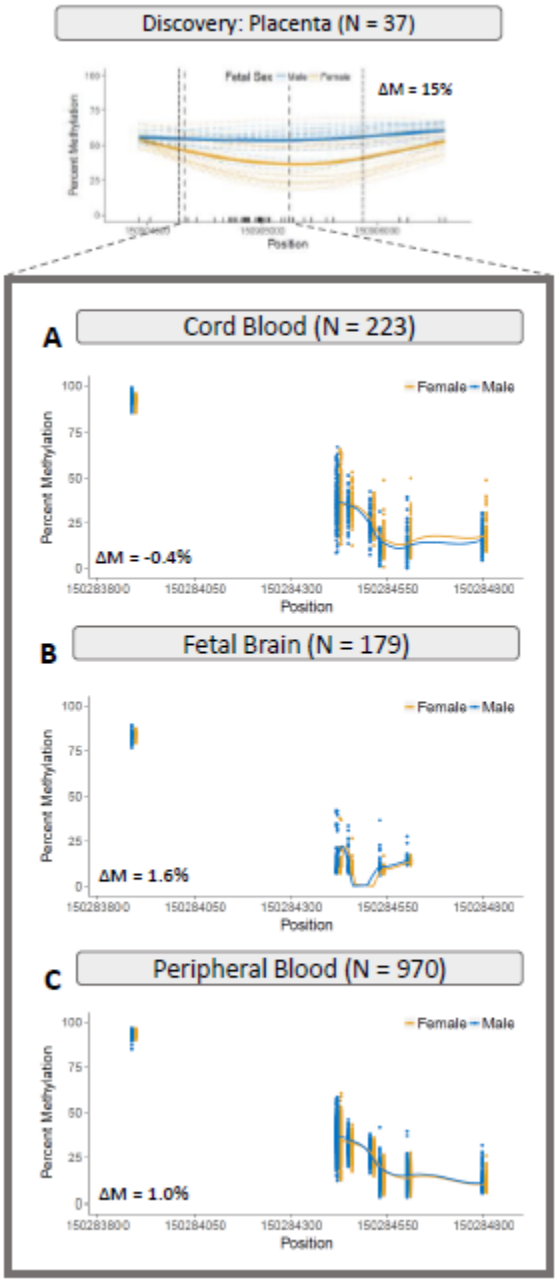



**Figure 5.3: Attempted replication of *ZNF300* DMR in additional tissue types. Panel A)** Cord blood samples from EARLI study **Panel B)** Fetal brain **Panel C)** Peripheral blood from 3-5 year old children.

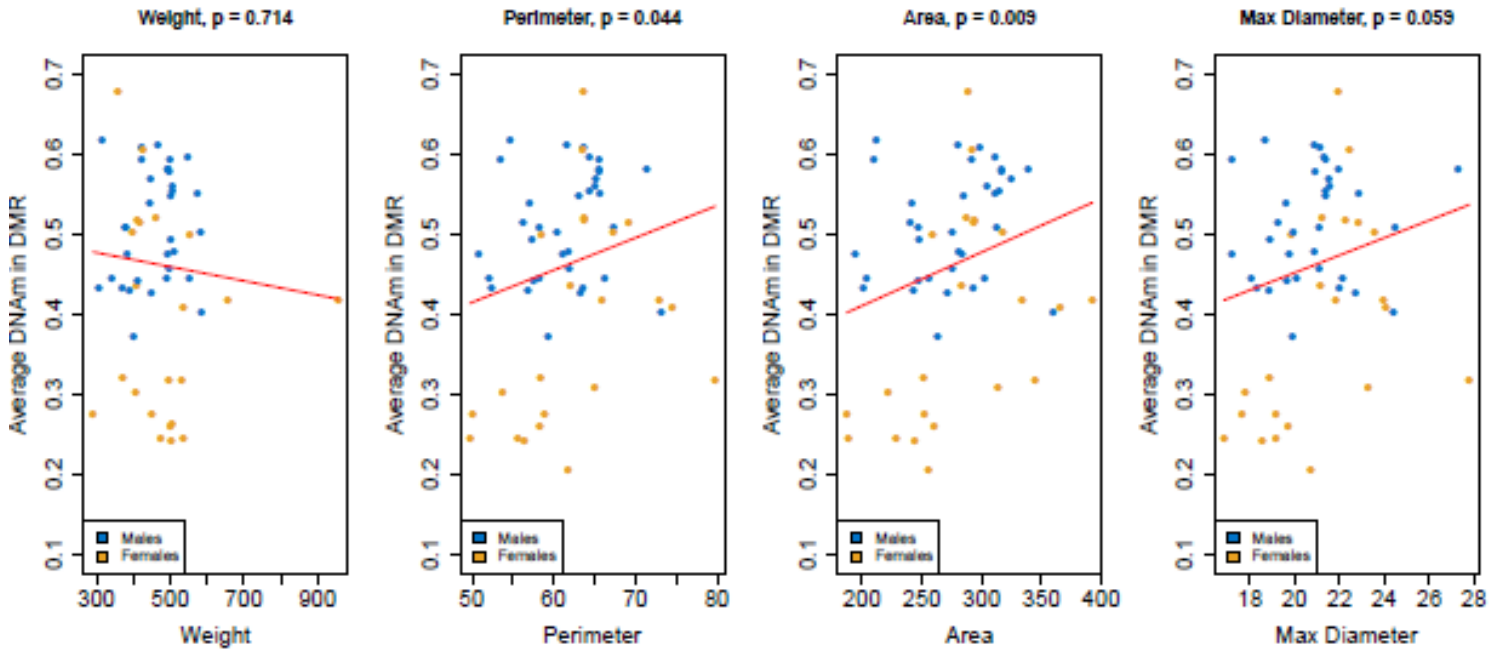



**Figure 5.4: Relationship between mean methylation levels at *ZNF300* DMR and placenta morphological features.** Mean DNAm levels in DMR (y-axis) vs. placental weight, perimeter, area, and maximum diameter (left to right, x-axis). P-values shown are for the regression of average DNAm in DMR onto each feature, adjusted for gestational age and birth weight.

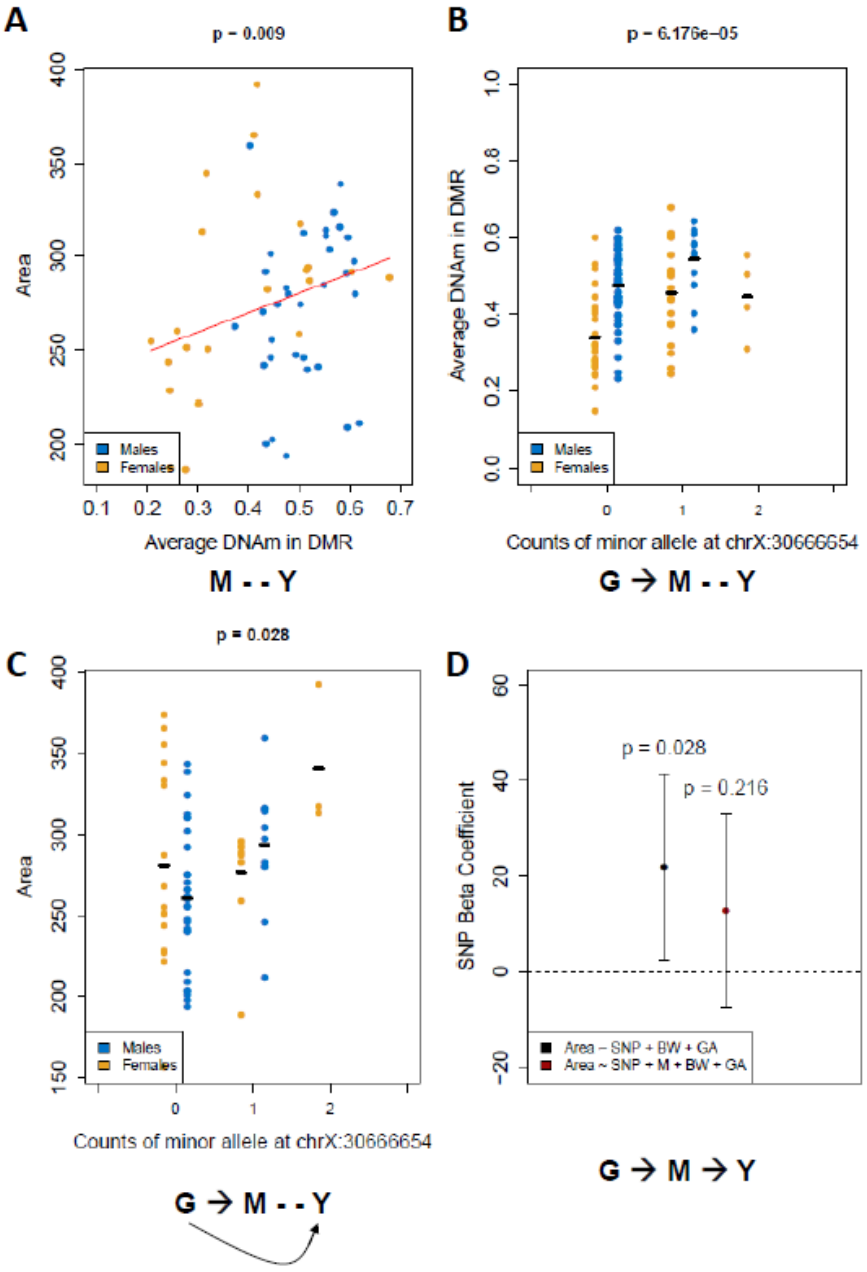



**Figure 5.5: Mediation analysis demonstrating relationship between a chromosome X SNP, mean *ZNF300* DMR methylation levels, and placenta area. Panel A)** Association between placenta area and average DNAm in DMR (p = 0.009); causality cannot be established. **Panel B)** Causal relationship between chromosome X SNP and average DNAm in DMR (p = 6.18E-5). **Panel C)** Causal relationship between chromosome X SNP and placenta area (p = 0.028). **Panel D)** Attenuation of SNP linear regression beta coefficient upon inclusion of DMR methylation level into model. All p-values shown for models that included birthweight and gestational age as covariates.

# CHAPTER 6: GENERAL DISCUSSION AND CONCLUDING REMARKS

## 6.1 Major Findings

The work in this dissertation provides mechanistic insights into autism spectrum disorder (ASD) and sexual dimorphism during pregnancy that can be gained through epigenetic data, in particular its integration with genotype data, across tissue types.

First, we integrated methylation quantitative trait loci (meQTL) information across fetal brain, peripheral blood, and cord blood with a recent iteration of GWAS summary statistics from the ASD subgroup of the Psychiatric Genomics Consortium (PGC-AUT) to better understand ASD genetic signals. We found that ASD-related SNPs were enriched for fetal brain (OR = 3.55, $P_{enrichment} < 0.001$) and peripheral blood meQTLs (OR = 1.58, $P_{enrichment} < 0.001$). Additionally, the CpG sites associated with ASD-related SNPs were preferentially engaged in immune response pathways across fetal brain, cord blood, and peripheral blood. This finding is particularly interesting given that it is consistent with previous gene expression studies of ASD but that it reveals pathways that are not implicated by ASD genetic findings. Finally, we showed, in general, how mapping of the CpG targets of ASD-associated SNPs can expand the biological insights gleaned through GWAS analysis.

Next, we focused on ASD studies of DNA methylation data from the placenta, which is an understudied but key tissue of interest in an ASD context. First, we repeated our previous integrative genotype-DNAm studies to include placenta meQTLs. Using a new iteration of the GWAS summary statistics from the PGC-AUT, we found ASD-related SNPs to be enriched for placenta meQTLs (fold enrichment = 1.25, $p_{Enrichment} = 0.011$), replicated our previous findings of enrichment for fetal brain (FE = 4.53, $p_{Enrichment} < 0.001$) and peripheral blood (FE = 1.39,

$p_{Enrichment}$ = 0.011) meQTLs, and also demonstrated enrichment for cord blood (FE = 2.16, $p_{Enrichment}$ < 0.001) meQTLs. While we did not replicate our previously observed enrichment for immune response pathways in ASD SNP-associated CpG sites across tissue types, tissue-specific pathways revealed interesting associations. These included Wnt signaling in peripheral blood, glial cell development in fetal brain, and virus-host interaction and syncytiotrophoblast cell development in placenta. Finally, all five genome-wide significant SNP loci identified by the PGC-AUT display evidence for meQTLs in placenta and/or peripheral blood, which expand loci to include additional gene targets and shed light on the tissue-specific functionality of these SNPs.

Our second investigation of placenta DNAm in an ASD context involved queries for differentially methylated regions (DMRs) and global methylation differences related to ASD diagnosis. In sex-stratified analyses, no significant DMRs were identified, though we did find suggestive evidence that global placenta DNAm patterns in children who went on to be classified as non-typically developing differed from those of children who went on to be classified as typically developing in CpG island shores (p = 0.043), CpG island shelves (p = 0.037), and open sea regions (p = 0.014). However, these differences were small in magnitude and did not remain significant after multiple testing correction.

Lastly we sought to discover fetal-sex-related differences in placenta DNAm, in an effort to contribute to research seeking to understand the mechanisms of sexual dimorphism during the gestational period. A CpG-island promoter region of *ZNF300* was found be approximately 15% hypermethylated in male placentas compared to female placentas. *ZNF300* has been previously associated with cell proliferation (Cai et al., 2014) and tumorogenesis (Wang et al., 2012). This

sex-related differentially methylated region (DMR) is placenta-specific, and replicated in 7 additional groups of samples, including 6 cohorts independent from the discovery sample. Replication sets demonstrated the persistence of this DMR at different points in the gestational period, in a single placenta cell type (villous cytotrophoblasts) and in different disease states (preeclampsia, intrauterine growth restriction, etc.) Female samples in the DMR region were characterized by two groups, one that exhibited DNAm levels similar to males, and the other that exhibited lower DNAm levels and who drives the differential methylation result. This later group was also characterized by smaller placenta area, perimeter, and maximum diameters as determined via 2D fetal surface imaging. Finally, suggestive evidence indicates that these methylation to placenta morphology relationships actually mediate local (chromosome 5) and distant (chromosome X) genetic relationships on placenta morphology that act through *ZNF300* DNAm.

These studies demonstrate the value of epigenetic data for insights into normal development and adverse health outcomes in early life, particularly when used in conjunction with paired genotype data. Moreover, the inherent nature of epigenetic marks to vary across tissue types was exploited to illustrate useful mechanisms that both span across tissues and that are tissue-specific.

### 6.2 Strengths and Limitations

The first two studies comprise the first genome-wide integration of genetic and epigenetic data in an ASD context that has been performed to date. While our studies are inspired by similar integrative studies that have been done for bipolar disorder and schizophrenia, our work is distinguished by the focus on both SNP-based and CpG-based enrichment analyses. Previous studies focused on the former analyses to demonstrate an enrichment of meQTLs in disorder-

associated SNPs. Our studies additionally enumerated the biological pathways engaged by ASD-controlled CpG sites, demonstrating additional value to be gained from genotype-epigenotype integration.

Another unique strength of the first two studies was the use of DNAm across tissue types. This approach treats a commonly considered challenge to studying epigenetic marks in life – that they inherently vary by tissue type – as an opportunity to gain both cross-tissue and tissue-specific insights. For example, in the first study, the CpG-based enrichment analyses revealed immune response pathways to be engaged by ASD SNPs through DNAm across fetal brain, cord, blood, and peripheral blood. In the second study, though many cross-tissue pathways were not discovered, the use of multiple tissue types allowed for tissue-specific patterns to be elucidated. In peripheral blood, Wnt signaling pathways were specifically enriched in CpG sites controlled by ASD SNPs. But this enrichment was not seen in cord blood, fetal brain, and placenta. The use of multiple tissue types additionally added strength to our SNP-based enrichment analyses by allowing for a rigorously designed comparison in which our SNP enrichment approach could be tested in a positive control tissue (fetal brain in both studies) and a negative control tissue (lung in the first study and adipose tissue in the second study). The results in these control tissues confirmed the validity of our methodological pipeline and defined a spectrum of results against which our placenta and blood tissues of interest could be evaluated.

They key insight from these two analyses is that all conclusions made from these studies are ones that are only achieved through the integration of genetic and epigenetic data, and not possible to glean by looking at either form of data in isolation. The SNP and CpG-based pipelines

demonstrated in these studies can be readily used in the context of any phenotype and tissue, as they only require a list of GWAS summary statistics and meQTL maps for the tissue of interest.

These cross-tissue meQTL studies did present inherent challenges. Most previous meQTL studies only analyzed a single tissue type from a single study; harmonizing specific meQTL and CpG targets across tissue types and across different studies required deeper attention to SNP and methylation measurement tools used, thresholds for calling a meQTL, and available power. Factors such as available SNPs to be tested, available CpG sites to be tested, and meQTL detection windows in meQTL discovery can all create differential opportunity for a SNP to be called a meQTL and a CpG site to be deemed as associated with a SNP. For example, placenta DNAm was measured via whole genome bisulfite sequencing (WGBS), which measures upwards of 28 million CpG sites in the human methylome. In contrast, meQTL lists for all other tissues were generated using DNAm measured via the Illumina 450K Array, which queries only 480,000 CpG sites in the methylome. Even within 450K studies, differences between tissues can exist in terms of SNP minor allele frequency and CpG site variability thresholds for inclusion into the meQTL query, quality control procedures, etc. These differences dictate that a SNP truly being a meQTL in one tissue and not another cannot be distinguished from a SNP only having an opportunity to be statistically classified as a meQTL in one tissue, but not the other. These concerns apply to classifying a CpG site as SNP-controlled in multiple tissues as well.

While the main purpose of these studies was to integrate genotype and DNA data, the fetal sex study was primarily a DNAm-only analysis, one that was characterized by the use of a comprehensive, gold standard assessment of the placenta methylome via WGBS. The chief strength of the fetal sex placenta methylation findings is the extent of replication for the

discovered *ZNF300* DMR. Overall the DMR persists with a consistent direction and similar magnitude in seven additional groups of samples, including an independent set of samples from the same Early Autism Risk Longitudinal Investigation, and six others from separate publically available datasets. These replication sets additionally demonstrate the robustness of the finding to different disease states, points in the gestational period, and potential confounding via cell type heterogeneity (Jaffe & Irizarry, 2014). Taken together, these facts dictate that this *ZNF300* DMR in placenta is the strongest evidence achieved to data for an autosomal DNAm difference related to sex, across any tissue. This study also strongly characterizes the nature and sources of this DMR by incorporating additional forms of data. For example, lookup of the *ZNF300* region in additional tissue types (including cord blood from the same individuals) allowed for the DMR to be classified as placenta-specific. Incorporation of placenta morphology data allowed for the downstream consequences of this DMR to begin to be elucidated and increased insight into how this DMR could mechanistically contribute to male-female differences during gestation. Finally, incorporation of paired genotype data allowed for insights into how the DMR could be genetically programed by local and distant SNP effects, and how *ZNF300* DNAm mediates these effects on placenta morphology.

This fetal sex study does suffer from a small sample size in the initial discovery phase (N = 37). This number dictated that the study could not adequately take advantage of the unique presence of WGBS-measured DNAm and was likely underpowered for association analysis. This study also did not employ direct correction for cell type heterogeneity given the lack of a current placenta cell type DNAm reference panel as has been developed for a variety of other tissues (Bakulski et al., 2016; Guintivano, Aryee, & Kaminsky, 2013; Reinius et al., 2012; Teschendorff

et al., 2016), and the lack of many methodological approaches for reference-free based correction in sequencing-based DNAm (Jaffe & Irizarry, 2014; Teschendorff & Zheng, 2017). It is unlikely that cell type heterogeneity drove the *ZNF300* finding given that it was unrelated to gestational age (of which cell type is a known surrogate) and the DMR was present when looking at DNAm from a single placenta cell type. However, lack of correction for cell type heterogeneity could have further underpowered our analyses and prevented us from discovering additional genome-wide significant loci (Teschendorff & Zheng, 2017). The limitations also apply to our WGBS study searching for DMRs related to ASD diagnosis.

Another limitation of the fetal sex study is the lack of direct evidence for the discovered DMR to influence *ZNF300* gene expression. Functional work has been performed to demonstrate the (expected) inverse association between *ZNF300* promoter methylation levels and corresponding gene expression levels, but this work was done in liver tissue, not placenta. Genotype-DNAm and genotype-DNAm-morphology studies also suffered from a key challenge in data integration efforts – available samples with all forms of data. In the analysis looking for genetic drivers of *ZNF300* methylation levels, only 121 samples were available, a number that left us underpowered to detect typical GWAS-level effect sizes. By contrast, the most recent PGC-AUT GWAS results were generated from 10,000 to 20,000 individuals (Grove et al., 2017). These concerns were exacerbated when limited to samples on which placenta morphology data were available as well.

### 6.3 Public Health Significance

The investigations integrating genetic and epigenetic data in an ASD context are of primary value for their insights into ASD etiology. The SNP and CpG-based enrichment studies are

global analyses that argue for the value of specific tissues in ASD epigenetic research. In the first study, peripheral blood shared the same SNP and CpG-based inferences as fetal brain, albeit at different effect sizes. This fact argues for blood-based epigenetic research as a surrogate for less accessible brain-based results for ASD. This conclusion should be limited to methylation signatures under some genetic control, since this was the focus of our study. But it may offer an exciting opportunity to investigate DNAm in an ASD context at scale, given that peripheral blood DNAm can be collected on very large groups of people.

Interestingly, in the second study, while the SNP-based enrichment results showed blood-brain concordance as we observed previously, CpG-based enrichment results pointed to tissue-specific pathways. This reopens the question about whether peripheral epigenetic signals can serve as a proxy for brain, but does indicate value in understanding how a peripheral tissue is contributing to or reflective of disease etiology. Indeed, if the timing of DNAm collection is after disease onset (as it was in these studies for peripheral blood), these relationships can be indicative of direct genotype to disease relationships, which itself manifests in epigenetic changes (Ladd-Acosta & Fallin, 2016).

For more affected tissues where DNAm collection precedes ASD onset, such as fetal brain and placenta, our genotype-epigenotype analyses primarily illuminated genotype to disease relationships mediated by DNAm. Particularly in the second study, our analyses illuminated mechanistic relationships linking specific (genome-wide significant) ASD SNPs to specific CpG sites and genes, identifying possible intervention targets and providing mechanistic insights into the functions of GWAS-associated SNPs (Ladd-Acosta & Fallin, 2016). Of particular interest in this latter endeavor is the function of GWAS SNPs across various tissues, which can offer

insights both if functions are common across tissues and if they are tissue-specific. The insights offered by GWAS can be enhanced if we consider their potential impact in a variety of contexts, starting with tissue type but also including factors such as specific windows in the life course and the presence of certain environmental exposures.

The fetal sex studies offer etiologic insights as well, potentially across a broad spectrum of phenotypes with sex-biased prevalence that occur across the life course. These include early life phenotypes such as preeclampsia and intrauterine growth restriction (Vatten & Skjaerven, 2004), but also later in life phenotypes like ASD (Christensen et al., 2016), schizophrenia (Aleman, Kahn, & Selten, 2003), cardiovascular outcomes (Bots, Peters, & Woodward, 2017), and autoimmune disorders (Ngo, Steyn, & McCombe, 2014). Our result, showing a fetal sex-related methylation difference related to placenta morphology, offers potential clues towards the developmental origins of this sexual dimorphism, as governed by the placenta.

Female samples in the identified DMR region were characterized by two groups: those that had methylation levels similar to males, and those had lower methylation levels. This latter group drove the differential methylation result, and also exhibited placentas that were smaller in area, perimeter, and maximum diameter. Though future work is necessary, it is possible that these two groups of females differ significantly in terms of their prognosis for early and later in life adverse health outcomes. Therefore, DNAm levels at this *ZNF300* region could be used to identify at risk populations and potentially inform treatment options. For example, if this dichotomy of females was found to correlate to adverse neurodevelopmental outcomes, placenta DNAm levels at *ZNF300* could guide early interventions.

**6.4 Future Directions**

A key area of future research for our studies integrating genetic and epigenetic data in an ASD context is to repeat analyses as ASD GWAS results are increasingly refined, and as meQTL information for additional tissue types becomes available. It is with this basis that the second study of this dissertation was developed; this work incorporated a more recent iteration of the GWAS summary statistics from the PGC-AUT, and an additional tissue of interest, the placenta. This iteration offered five genome-wide significant loci and therefore was of most value in the GWAS loci expansion analysis, which demonstrated meQTL relationships for all of these SNPs. Functional follow-up of these SNPs to understand their mechanistic role in ASD is likely to be undertaken, and these investigations can be informed by the nature of genetic control of DNAm that these SNPs exhibit in placenta and peripheral blood.

As the field of epigenetic epidemiology advances, it will be important to gain additional coverage of the DNAm methylome in large groups of samples across different tissue types. We discovered that all five genome-wide significant loci identified in the PGC-AUT displayed evidence for controlling nearby placenta DNAm levels, which were measured via WGBS. This is likely not a consequence of a unique engagement by the placenta methylome of ASD SNPs, but rather a function of the significantly greater opportunity of SNPs to be called a placenta meQTL. Therefore, WGBS-level DNAm in additional tissues such as fetal brain, peripheral blood, and cord blood, can offer both comprehensive assessment of meQTLs and more unbiased cross-tissue comparison. If possible, meQTL queries should be done on multiple tissue samples from the same group of individuals, using identical parameters for the meQTL queries. These efforts can minimize common issues that complicate comparing meQTLs identified in different studies. Increased availability of WGBS-level DNAm can also aid in "methylation-only" studies to

discover DMRs and global methylation differences related to ASD; our placenta WGBS ASD study suffered from small sample size because we conducted sex-stratified analyses.

Discovery of the fetal sex *ZNF300* DMR in placenta offers multiple avenues for future research. Given the extent of replication that was observed, and future efforts should focus on identifying sources of the DMR and its downstream morphological and phenotypic consequences. First, *ZNF300* promoter DNAm levels should be confirmed for their role in dictating *ZNF300* expression. While an inverse relationship is canonical and has been demonstrated in liver cells, rigorous functional work in placenta should be established. Next, additional samples must be generated on which genotype, DNAm, and placenta morphology data all are available. These data will be useful to better understand the interplay of these factors in establishing the suggestive genotype to *ZNF300* DNAm to placenta morphology relationship discovered herein. Moreover, it will be increasingly necessary to integrate these data to understand placenta function more broadly.

Other follow up efforts to the *ZNF300* finding must focus on the observed bimodality in female samples. First, this clustering of females samples should be interpreted as differential variability compared to male samples, not simply as a sex-specific differences in mean values. It is possible that the female placenta methylome is broadly characterized as more variable than that of males. Therefore, methodologies should be developed to identify differentially variable regions in sequencing-based DNAm data and should be applied in this context. Currently, such methods are optimized for DNA methylation array data. Next, the sources of the *ZNF300* female bimodality should be enumerated. It is possible that one group of female samples shared a common environmental insult during pregnancy, or certain maternal characteristics. These possibilities

should be investigated further, though they require a unique study setting in which both placenta DNAm and rich maternal exposure and/or behavior history during pregnancy is available. Finally, it is of great interest to characterize the effects of lower *ZNF300* methylation levels in females, and their subsequent decreased placenta areas, on health outcomes in pregnancy and over the life course. Such investigations will require paired DNAm and placenta morphology data, along with potentially long term prospective follow up.

The work in this dissertation provides demonstrates the value of epigenetic factors in discovering insights into ASD and sexual dimorphism during pregnancy. These findings were both enhanced and characterized in key ways by using many techniques to incorporate genetic data in conjunction with epigenetic data, and by exploiting DNAm data from a variety of tissue types.

### 6.5 References

Aleman, A., Kahn, R. S., & Selten, J.-P. (2003). Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Archives of General Psychiatry*, *60*(6), 565–571. https://doi.org/10.1001/archpsyc.60.6.565

Bakulski, K. M., Feinberg, J. I., Andrews, S. V., Yang, J., Brown, S., L McKenney, S., … Fallin, M. D. (2016). DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics*, *11*(5), 354–362. https://doi.org/10.1080/15592294.2016.1161875

Bots, S. H., Peters, S. A. E., & Woodward, M. (2017). Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health*, *2*(2), e000298. https://doi.org/10.1136/bmjgh-2017-000298

Cai, J., Gong, R., Yan, F., Yu, C., Liu, L., Wang, W., … Huang, Z. (2014). ZNF300 knockdown inhibits forced megakaryocytic differentiation by phorbol and erythrocytic differentiation by arabinofuranosyl cytidine in K562 cells. *PloS One*, *9*(12), e114768. https://doi.org/10.1371/journal.pone.0114768

Christensen, D. L., Baio, J., Van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Centers for Disease Control and Prevention (CDC). (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years--Autism and Developmental

Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)*, *65*(3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R., Won, H., … Børglum, A. D. (2017). Common risk variants identified in autism spectrum disorder. *bioRxiv*, 224774. https://doi.org/10.1101/224774

Guintivano, J., Aryee, M. J., & Kaminsky, Z. A. (2013). A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics*, *8*(3), 290–302. https://doi.org/10.4161/epi.23924

Jaffe, A. E., & Irizarry, R. A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology*, *15*(2), R31. https://doi.org/10.1186/gb-2014-15-2-r31

Ladd-Acosta, C., & Fallin, M. D. (2016). The role of epigenetics in genetic and environmental epidemiology. *Epigenomics*, *8*(2), 271–283. https://doi.org/10.2217/epi.15.102

Ngo, S. T., Steyn, F. J., & McCombe, P. A. (2014). Gender differences in autoimmune disease. *Frontiers in Neuroendocrinology*, *35*(3), 347–369. https://doi.org/10.1016/j.yfrne.2014.04.004

Reinius, L. E., Acevedo, N., Joerink, M., Pershagen, G., Dahlén, S.-E., Greco, D., … Kere, J. (2012). Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PloS One*, *7*(7), e41361. https://doi.org/10.1371/journal.pone.0041361

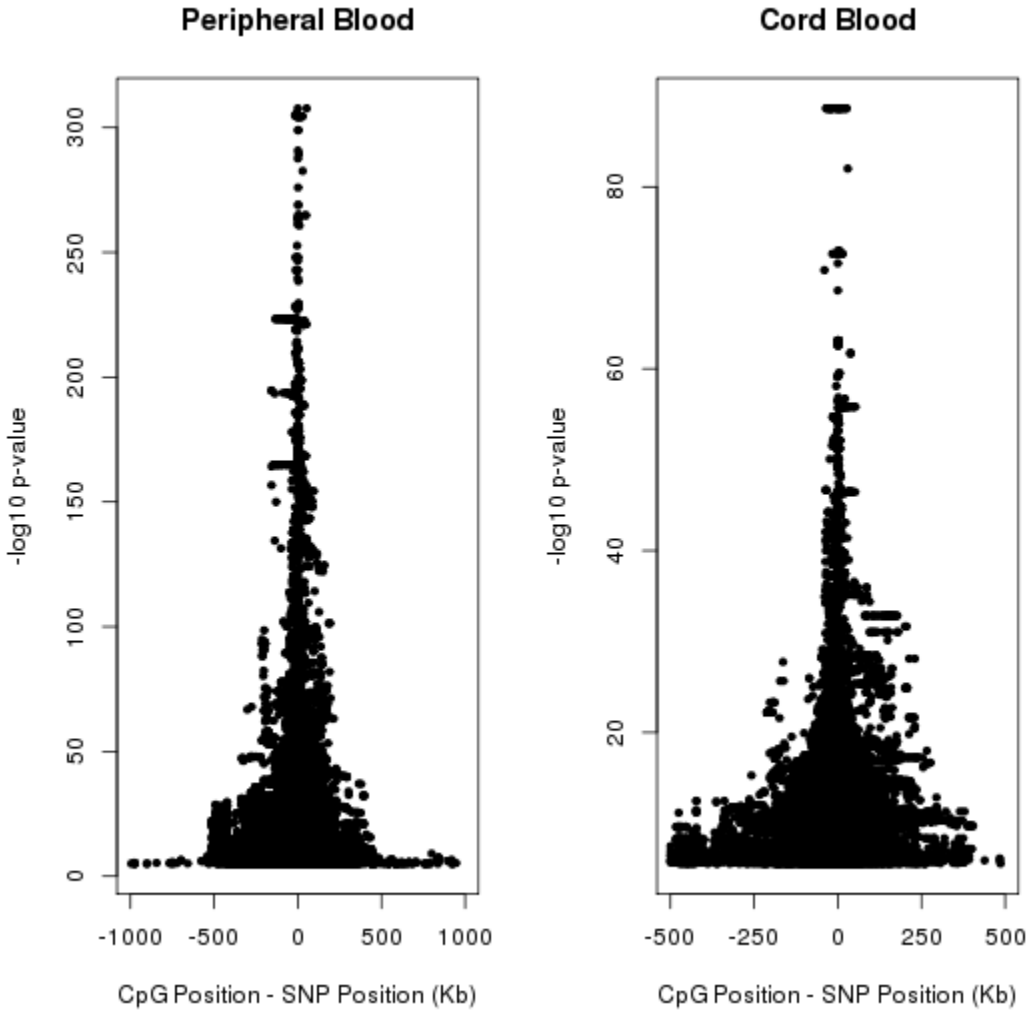
Teschendorff, A. E., Gao, Y., Jones, A., Ruebner, M., Beckmann, M. W., Wachter, D. L., … Widschwendter, M. (2016). DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nature Communications*, *7*, 10478. https://doi.org/10.1038/ncomms10478

Teschendorff, A. E., & Zheng, S. C. (2017). Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics*, *9*(5), 757–768. https://doi.org/10.2217/epi-2016-0153

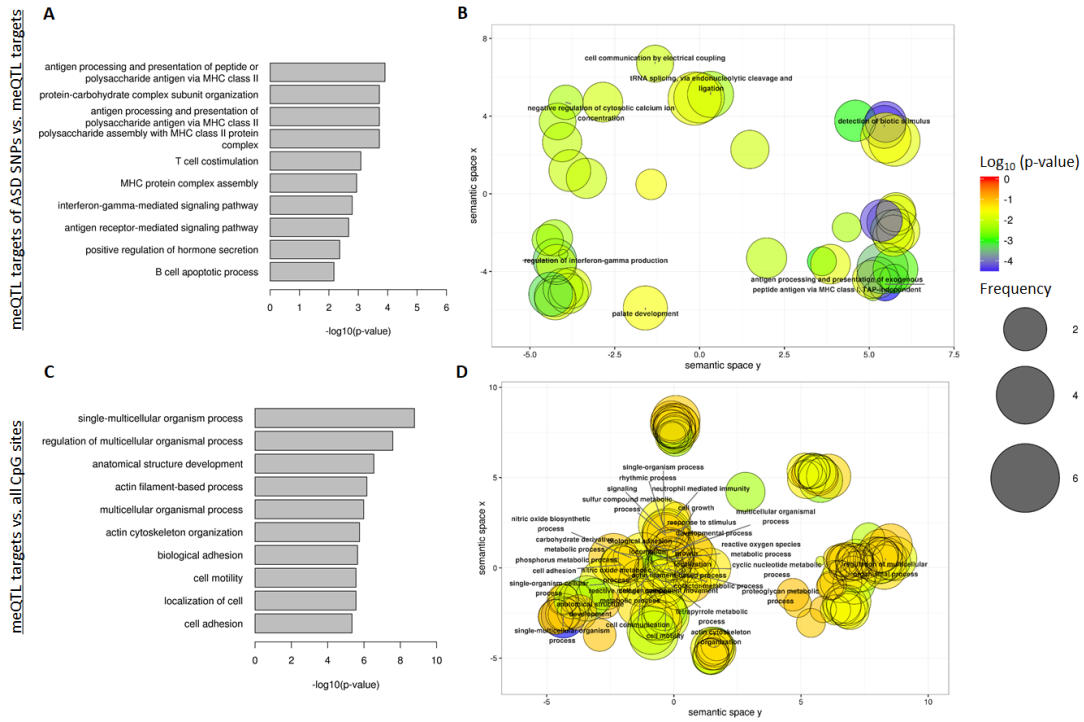
Vatten, L. J., & Skjaerven, R. (2004). Offspring sex and pregnancy outcome by length of gestation. *Early Human Development*, *76*(1), 47–54.

Wang, T., Wang, X., Xu, J., Wu, X.-P., Qiu, H., Yi, H., & Li, W.-X. (2012). Overexpression of the human ZNF300 gene enhances growth and metastasis of cancer cells through activating NF-kB pathway. *Journal of Cellular and Molecular Medicine*, *16*(5), 1134–1145. https://doi.org/10.1111/j.1582-4934.2011.01388.x

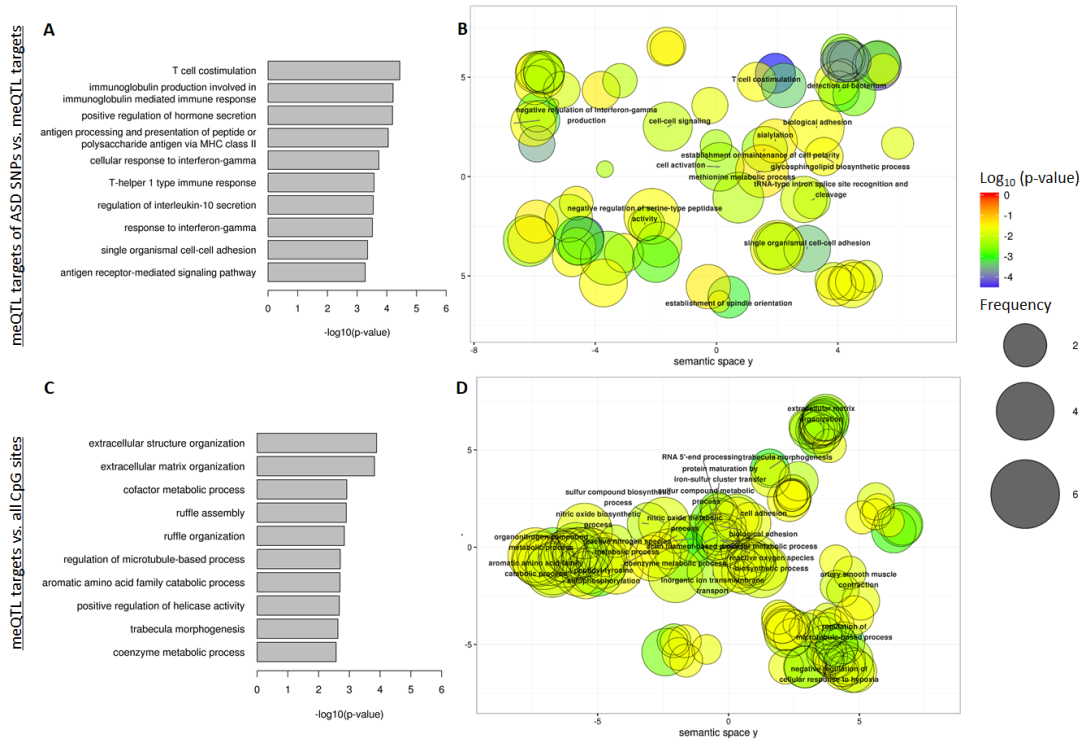
# APPENDIX I



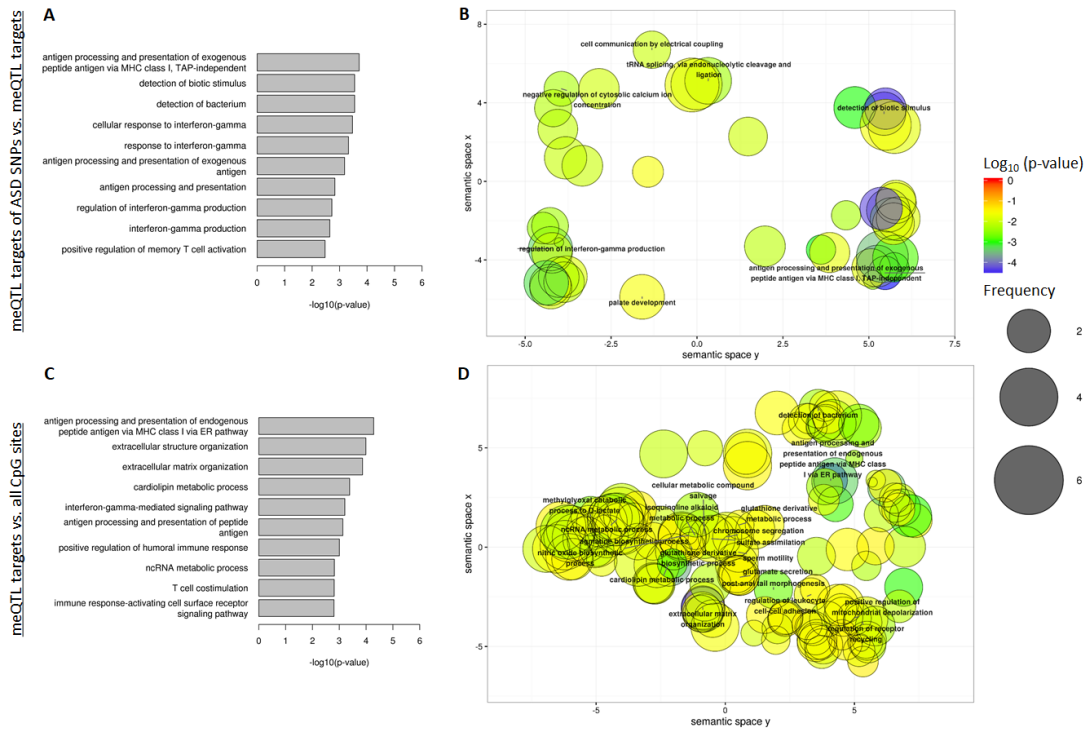

**Supplementary Figure I-1**: **The relationship between degree of significance and distance between SNP and CpG site on chromosome 21.** Degree of significance (y-axis) defined by -log10 p-value. Only meQTLs present at FDR = 5% are shown. The degree of significance decays with increasing distance. Left panel shows relationship for SEED peripheral blood data and right panel shows relationship for EARLI cord blood data.




**Supplementary Figure I-2**: **Gene Ontology enrichment analysis for meQTL targets in peripheral blood.** Gene Ontology (GO) enrichment analysis via the 'gometh' function in the *MissMethyl* R package. The results for the "biological processes" group were pruned to remove overlapping terms using the REVIGO software. Results shown for meQTL targets of ASD-related (PGC p-value < 1E-4) SNPs and their proxies ($r^2$) against a background of all meQTL targets (n = 201 vs n = 59,308; **Panels A+B**) and for meQTL targets against a background of all CpG sties tested (n = 59,308 vs n = 290,066; **Panels C+D**). meQTL targets all defined via meQTL p-value threshold = FDR 5%. **Panels A+C**) The top 10 biological process by GO enrichment p-value after REVIGO pruning. **Panel B+D**) A multi-dimensional scaling projection of the semantic similarity in nominally significant (enrichment p-value < 0.05) GO terms produced by REVIGO. Clusters are identified via labeling of the terms with both the least redundancy and highest degree of enrichment ('dispensability' value < 0.15). Color reflect degree of significance and increasing size reflects greater frequency of term in GO database.




**Supplementary Figure I-3**: **Gene Ontology enrichment analysis for meQTL targets in cord blood.** Gene Ontology (GO) enrichment analysis via the 'gometh' function in the *MissMethyl* R package. The results for the "biological processes" group were pruned to remove overlapping terms using the REVIGO software. Results shown for meQTL targets of ASD-related (PGC p-value < 1E-4) SNPs and their proxies ($r^2$) against a background of all meQTL targets (n = 66 vs n = 22,803; **Panels A+B**) and for meQTL targets against a background of all CpG sties tested (n = 22,803 vs n = 289,645; **Panels C+D**). meQTL targets all defined via meQTL p-value threshold = FDR 5%. **Panels A+C**) The top 10 biological process by GO enrichment p-value after REVIGO pruning. **Panel B+D**) A multi-dimensional scaling projection of the semantic similarity in nominally significant (enrichment p-value < 0.05) GO terms produced by REVIGO. Clusters are identified via labeling of the terms with both the least redundancy and highest degree of enrichment ('dispensability' value < 0.15). Color reflect degree of significance and increasing size reflects greater frequency of term in GO database.




**Supplementary Figure I-4**: **Gene Ontology enrichment analysis for meQTL targets in fetal brain.** Gene Ontology (GO) enrichment analysis via the 'gometh' function in the *MissMethyl* R package. The results for the "biological processes" group were pruned to remove overlapping terms using the REVIGO software. Results shown for meQTL targets of ASD-related (PGC p-value < 1E-4) SNPs and their proxies ($r^2$) against a background of all meQTL targets (n = 53 vs n = 7,863; **Panels A+B**) and for meQTL targets against a background of all CpG sties tested (n = 7,863 vs n = 314,554; **Panels C+D**). meQTL targets all defined via meQTL p-value threshold = FDR 5%. **Panels A+C**) The top 10 biological process by GO enrichment p-value after REVIGO pruning. **Panel B+D**) A multi-dimensional scaling projection of the semantic similarity in nominally significant (enrichment p-value < 0.05) GO terms produced by REVIGO. Clusters are identified via labeling of the terms with both the least redundancy and highest degree of enrichment ('dispensability' value < 0.15). Color reflect degree of significance and increasing size reflects greater frequency of term in GO database.

**Supplementary Table I-1: Number of genes implicated by meQTL results across tissue type**

| | *Number of genes with meQTLs* | *Number of genes with meQTL targets* | *Number of non-overlapping genes among meQTL targets* |
|---|---|---|---|
| **Fetal Brain** | 7,199[a] | 2,784[a] | 144[a] |
| **Peripheral Blood** | 18,454[b]<br>18,232[c]<br>17,655[d] | 13,244[b]<br>12,483[c]<br>11,079[d] | 597[b]<br>541[c]<br>469[d] |
| **Cord Blood** | 15,869[b]<br>15,254[c]<br>14,015[d] | 8,647[b]<br>7,876[c]<br>6,700[d] | 472[b]<br>352[c]<br>226[d] |
| **Lung** | 7,487[c] | 8,087[c] | 2,846[c] |

[a]FDR not specified. [b]FDR = 10% [c]FDR = 5% [d]FDR = 1%

**Supplementary Table I-2: Summary of meQTL evidence across tissue type.**

| Scenario | Blood | Cord Blood | Fetal Brain | SNPs | % of Total SNPs | Independent Sites | % of Total Independent Sites |
|---|---|---|---|---|---|---|---|
| ***1*** | ✓ | ✓ | ✓ | 125,869 | 4.65% | 6,640 | 4.15% |
| ***2*** | ✓ | ✓ | × | 407,722 | 15.08% | 22,135 | 13.83% |
| ***3*** | ✓ | × | ✓ | 30,691 | 1.14% | 1,354 | 0.85% |
| ***4*** | ✓ | × | × | 722,703 | 26.73% | 42,561 | 26.58% |
| ***5*** | × | ✓ | ✓ | 528 | 0.02% | 18 | 0.01% |
| ***6*** | × | ✓ | × | 6,299 | 0.23% | 333 | 0.21% |
| ***7*** | × | × | ✓ | 4,940 | 0.18% | 237 | 0.15% |
| ***8*** | × | × | × | 1,405,261 | 51.97% | 86,821 | 54.23% |
| | | | **SUM** | 2,704,013 | | 160,099 | |

Results are shown for meQTLs associated at FDR = 5% threshold in blood and cord blood datasets and threshold of 1E-8 in fetal brain.

Only SNPs that were included in all three tissues in their respective meQTL queries are included in this analyis (n = 2,704,013).

Independent sites were constructed by grouping SNPs into bins defined by recombination hot spot data from 1000 Genomes (see Methods).

For example, scenario 1 lists that there are a total of 125,869 SNPs that are meQTLs in blood, cord blood, and fetal brain, which fall into 6,640 loci.

**Supplementary Table I-3: Summary of meQTL evidence for ASD-related (p < 1e-04) PGC results.**

| Scenario | Blood | Cord Blood | Fetal Brain | SNPs | % of Total SNPs | Independent Sites | % of Total Independent Sites |
|---|---|---|---|---|---|---|---|
| ***1*** | ✓ | ✓ | ✓ | 5 | 0.46% | 2 | 0.80% |
| ***2*** | ✓ | ✓ | × | 74 | 6.76% | 18 | 7.23% |
| ***3*** | ✓ | × | ✓ | 0 | 0.00% | 0 | 0.00% |
| ***4*** | ✓ | × | × | 195 | 17.82% | 28 | 11.24% |
| ***5*** | × | ✓ | ✓ | 19 | 1.74% | 8 | 3.21% |
| ***6*** | × | ✓ | × | 75 | 6.86% | 13 | 5.22% |
| ***7*** | × | × | ✓ | 0 | 0.00% | 0 | 0.00% |
| ***8*** | × | × | × | 726 | 66.36% | 180 | 72.29% |
| | | | **SUM** | 1094 | | 249 | |

Results are shown for meQTLs associated at FDR = 5% threshold in blood and cord blood datasets and threshold of 1E-8 in fetal brain.

All ASD-related (p< 1E-04) SNPs in PGC, regardless of if they were tested in the respective meQTL studies, are included in this analysis.

Independent sites were constructed by grouping SNPs into bins defined by recombination hot spot data from 1000 Genomes (see Methods).

For example, scenario 1 lists that there are a total of 5 SNPs that are meQTLs in blood, cord blood, and fetal brain, which fall into 2 loci.

# APPENDIX II

**Supplementary Data II-1: Peripheral blood meQTLs identified at FDR = 5%**.

Note: This table is too large to be included in this dissertation but can be found at https://www.nature.com/articles/s41467-017-00868-y

**Supplementary Data II-2: Cord blood meQTLs identified at FDR = 5%**.

Note: This table is too large to be included in this dissertation but can be found at https://www.nature.com/articles/s41467-017-00868-y

**Supplementary Data II-3: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in peripheral blood**. P.DE refers to p-value for enrichment as computed by gometh() function in MissMethyl R package. Scaled Rank is computed as rank divided by the total number of terms in the list (n = 95).

| GO ID | Term | P.DE | Rank | ScaledRank |
|---|---|---|---|---|
| ***GO:0002504*** | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | 0.000123 | 1 | 0.01 |
| ***GO:0002505*** | antigen processing and presentation of polysaccharide antigen via MHC class II | 0.000192 | 2 | 0.02 |
| ***GO:0002506*** | polysaccharide assembly with MHC class II protein complex | 0.000192 | 3 | 0.03 |
| ***GO:0071823*** | protein-carbohydrate complex subunit organization | 0.000192 | 4 | 0.04 |
| ***GO:0031295*** | T cell costimulation | 0.000816 | 5 | 0.05 |
| ***GO:0002396*** | MHC protein complex assembly | 0.001139 | 6 | 0.06 |
| ***GO:0060333*** | interferon-gamma-mediated signaling pathway | 0.001607 | 7 | 0.07 |
| ***GO:0050851*** | antigen receptor-mediated signaling pathway | 0.002121 | 8 | 0.08 |
| ***GO:0046887*** | positive regulation of hormone secretion | 0.0043 | 9 | 0.09 |
| ***GO:0001783*** | B cell apoptotic process | 0.006717 | 10 | 0.11 |
| ***GO:0030166*** | proteoglycan biosynthetic process | 0.006887 | 11 | 0.12 |
| ***GO:1902775*** | mitochondrial large ribosomal subunit assembly | 0.007342 | 12 | 0.13 |
| ***GO:0034341*** | response to interferon-gamma | 0.007598 | 13 | 0.14 |
| ***GO:0045338*** | farnesyl diphosphate metabolic process | 0.010001 | 14 | 0.15 |
| ***GO:1902856*** | negative regulation of nonmotile primary cilium assembly | 0.010681 | 15 | 0.16 |
| ***GO:0006369*** | termination of RNA polymerase II transcription | 0.010682 | 16 | 0.17 |
| ***GO:0032625*** | interleukin-21 production | 0.012311 | 17 | 0.18 |
| ***GO:0072619*** | interleukin-21 secretion | 0.012311 | 18 | 0.19 |
| ***GO:1901899*** | positive regulation of relaxation of cardiac muscle | 0.013317 | 19 | 0.2 |
| ***GO:1902570*** | protein localization to nucleolus | 0.013546 | 20 | 0.21 |
| ***GO:0031099*** | regeneration | 0.013974 | 21 | 0.22 |
| ***GO:0016337*** | single organismal cell-cell adhesion | 0.014138 | 22 | 0.23 |
| ***GO:0002381*** | immunoglobulin production involved in immunoglobulin mediated immune response | 0.016465 | 23 | 0.24 |

| | | | | |
|---|---|---|---|---|
| ***GO:0048024*** | regulation of mRNA splicing, via spliceosome | 0.016621 | 24 | 0.25 |
| ***GO:0002339*** | B cell selection | 0.016698 | 25 | 0.26 |
| ***GO:0061470*** | T follicular helper cell differentiation | 0.017096 | 26 | 0.27 |
| ***GO:0031635*** | adenylate cyclase-inhibiting opioid receptor signaling pathway | 0.018049 | 27 | 0.28 |
| ***GO:0051012*** | microtubule sliding | 0.019357 | 28 | 0.29 |
| ***GO:0044060*** | regulation of endocrine process | 0.019483 | 29 | 0.31 |
| ***GO:0042461*** | photoreceptor cell development | 0.021344 | 30 | 0.32 |
| ***GO:0045218*** | zonula adherens maintenance | 0.022584 | 31 | 0.33 |
| ***GO:0098602*** | single organism cell adhesion | 0.023398 | 32 | 0.34 |
| ***GO:0030299*** | intestinal cholesterol absorption | 0.023474 | 33 | 0.35 |
| ***GO:0048597*** | post-embryonic camera-type eye morphogenesis | 0.023496 | 34 | 0.36 |
| ***GO:0010899*** | regulation of phosphatidylcholine catabolic process | 0.023678 | 35 | 0.37 |
| ***GO:1900155*** | negative regulation of bone trabecula formation | 0.024242 | 36 | 0.38 |
| ***GO:0014717*** | regulation of satellite cell activation involved in skeletal muscle regeneration | 0.024316 | 37 | 0.39 |
| ***GO:0010046*** | response to mycotoxin | 0.025255 | 38 | 0.4 |
| ***GO:1902572*** | negative regulation of serine-type peptidase activity | 0.025549 | 39 | 0.41 |
| ***GO:1901509*** | regulation of endothelial tube morphogenesis | 0.026235 | 40 | 0.42 |
| ***GO:0030091*** | protein repair | 0.026842 | 41 | 0.43 |
| ***GO:0044272*** | sulfur compound biosynthetic process | 0.027027 | 42 | 0.44 |
| ***GO:0032844*** | regulation of homeostatic process | 0.027254 | 43 | 0.45 |
| ***GO:1903361*** | protein localization to basolateral plasma membrane | 0.027412 | 44 | 0.46 |
| ***GO:0097021*** | lymphocyte migration into lymphoid organs | 0.028288 | 45 | 0.47 |
| ***GO:1902571*** | regulation of serine-type peptidase activity | 0.028416 | 46 | 0.48 |
| ***GO:0006285*** | base-excision repair, AP site formation | 0.028604 | 47 | 0.49 |
| ***GO:0006029*** | proteoglycan metabolic process | 0.029195 | 48 | 0.51 |
| ***GO:0030263*** | apoptotic chromosome condensation | 0.029381 | 49 | 0.52 |
| ***GO:0042117*** | monocyte activation | 0.030469 | 50 | 0.53 |
| ***GO:0019557*** | histidine catabolic process to glutamate and formate | 0.030471 | 51 | 0.54 |

| | | | | |
|---|---|---|---|---|
| ***GO:0043606*** | formamide metabolic process | 0.030471 | 52 | 0.55 |
| ***GO:0032940*** | secretion by cell | 0.030778 | 53 | 0.56 |
| ***GO:0070544*** | histone H3-K36 demethylation | 0.031005 | 54 | 0.57 |
| ***GO:0016075*** | rRNA catabolic process | 0.031544 | 55 | 0.58 |
| ***GO:0048852*** | diencephalon morphogenesis | 0.032019 | 56 | 0.59 |
| ***GO:0061146*** | Peyer's patch morphogenesis | 0.032314 | 57 | 0.6 |
| ***GO:0036124*** | histone H3-K9 trimethylation | 0.032359 | 58 | 0.61 |
| ***GO:0042159*** | lipoprotein catabolic process | 0.03293 | 59 | 0.62 |
| ***GO:0036301*** | macrophage colony-stimulating factor production | 0.033104 | 60 | 0.63 |
| ***GO:1901256*** | regulation of macrophage colony-stimulating factor production | 0.033104 | 61 | 0.64 |
| ***GO:0002440*** | production of molecular mediator of immune response | 0.033722 | 62 | 0.65 |
| ***GO:0009223*** | pyrimidine deoxyribonucleotide catabolic process | 0.034189 | 63 | 0.66 |
| ***GO:0008635*** | activation of cysteine-type endopeptidase activity involved in apoptotic process by cytochrome c | 0.034482 | 64 | 0.67 |
| ***GO:0046349*** | amino sugar biosynthetic process | 0.034833 | 65 | 0.68 |
| ***GO:1901565*** | organonitrogen compound catabolic process | 0.036166 | 66 | 0.69 |
| ***GO:0006024*** | glycosaminoglycan biosynthetic process | 0.037186 | 67 | 0.71 |
| ***GO:0045471*** | response to ethanol | 0.037351 | 68 | 0.72 |
| ***GO:0031124*** | mRNA 3'-end processing | 0.037546 | 69 | 0.73 |
| ***GO:0006641*** | triglyceride metabolic process | 0.038112 | 70 | 0.74 |
| ***GO:0006987*** | activation of signaling protein activity involved in unfolded protein response | 0.038374 | 71 | 0.75 |
| ***GO:0010225*** | response to UV-C | 0.038784 | 72 | 0.76 |
| ***GO:0030202*** | heparin metabolic process | 0.038807 | 73 | 0.77 |
| ***GO:0030210*** | heparin biosynthetic process | 0.038807 | 74 | 0.78 |
| ***GO:1900745*** | positive regulation of p38MAPK cascade | 0.041265 | 75 | 0.79 |
| ***GO:0032471*** | negative regulation of endoplasmic reticulum calcium ion concentration | 0.041845 | 76 | 0.8 |
| ***GO:0006488*** | dolichol-linked oligosaccharide biosynthetic process | 0.042594 | 77 | 0.81 |
| ***GO:0006490*** | oligosaccharide-lipid intermediate biosynthetic process | 0.042594 | 78 | 0.82 |
| ***GO:0006244*** | pyrimidine nucleotide catabolic process | 0.043163 | 79 | 0.83 |

| | | | | |
|---|---|---|---|---|
| ***GO:0030521*** | androgen receptor signaling pathway | 0.043956 | 80 | 0.84 |
| ***GO:0002407*** | dendritic cell chemotaxis | 0.044505 | 81 | 0.85 |
| ***GO:0042118*** | endothelial cell activation | 0.044915 | 82 | 0.86 |
| ***GO:0000185*** | activation of MAPKKK activity | 0.045366 | 83 | 0.87 |
| ***GO:0014719*** | skeletal muscle satellite cell activation | 0.04705 | 84 | 0.88 |
| ***GO:0006353*** | DNA-templated transcription, termination | 0.047062 | 85 | 0.89 |
| ***GO:0051597*** | response to methylmercury | 0.047182 | 86 | 0.91 |
| ***GO:0032211*** | negative regulation of telomere maintenance via telomerase | 0.047797 | 87 | 0.92 |
| ***GO:0006638*** | neutral lipid metabolic process | 0.047806 | 88 | 0.93 |
| ***GO:0000027*** | ribosomal large subunit assembly | 0.047898 | 89 | 0.94 |
| ***GO:0009064*** | glutamine family amino acid metabolic process | 0.048245 | 90 | 0.95 |
| ***GO:0006002*** | fructose 6-phosphate metabolic process | 0.048597 | 91 | 0.96 |
| ***GO:0015942*** | formate metabolic process | 0.048828 | 92 | 0.97 |
| ***GO:1900121*** | negative regulation of receptor binding | 0.049665 | 93 | 0.98 |
| ***GO:0001704*** | formation of primary germ layer | 0.049838 | 94 | 0.99 |
| ***GO:0001547*** | antral ovarian follicle growth | 0.049941 | 95 | 1 |

**Supplementary Data II-4: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in cord blood**. P.DE refers to p-value for enrichment as computed by gometh() function in MissMethyl R package. Scaled Rank is computed as rank divided by the total number of terms in the list (n = 76).

| GO ID | Term | P.DE | Rank | ScaledRank |
|---|---|---|---|---|
| GO:0031295 | T cell costimulation | 3.66E-05 | 1 | 0.01 |
| GO:0002381 | immunoglobulin production involved in immunoglobulin mediated immune response | 6.16E-05 | 2 | 0.03 |
| GO:0046887 | positive regulation of hormone secretion | 6.42E-05 | 3 | 0.04 |
| GO:0002504 | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | 8.99E-05 | 4 | 0.05 |
| GO:0071346 | cellular response to interferon-gamma | 0.000185 | 5 | 0.07 |
| GO:0042088 | T-helper 1 type immune response | 0.000269 | 6 | 0.08 |
| GO:2001179 | regulation of interleukin-10 secretion | 0.000286 | 7 | 0.09 |
| GO:0034341 | response to interferon-gamma | 0.000305 | 8 | 0.11 |
| GO:0016337 | single organismal cell-cell adhesion | 0.000444 | 9 | 0.12 |
| GO:0050851 | antigen receptor-mediated signaling pathway | 0.000534 | 10 | 0.13 |
| GO:0032689 | negative regulation of interferon-gamma production | 0.000684 | 11 | 0.14 |
| GO:0098602 | single organism cell adhesion | 0.000803 | 12 | 0.16 |
| GO:0051294 | establishment of spindle orientation | 0.000914 | 13 | 0.17 |
| GO:0016045 | detection of bacterium | 0.001086 | 14 | 0.18 |
| GO:0051046 | regulation of secretion | 0.001438 | 15 | 0.2 |
| GO:0030155 | regulation of cell adhesion | 0.001603 | 16 | 0.21 |
| GO:0002440 | production of molecular mediator of immune response | 0.00192 | 17 | 0.22 |
| GO:0032673 | regulation of interleukin-4 production | 0.001953 | 18 | 0.24 |
| GO:0032613 | interleukin-10 production | 0.002202 | 19 | 0.25 |
| GO:0009595 | detection of biotic stimulus | 0.002516 | 20 | 0.26 |
| GO:0098609 | cell-cell adhesion | 0.002522 | 21 | 0.28 |
| GO:0032633 | interleukin-4 production | 0.002656 | 22 | 0.29 |
| GO:0046903 | secretion | 0.003056 | 23 | 0.3 |
| GO:0002437 | inflammatory response to antigenic stimulus | 0.003388 | 24 | 0.32 |
| GO:0007163 | establishment or maintenance of cell polarity | 0.004132 | 25 | 0.33 |
| GO:0055082 | cellular chemical homeostasis | 0.00453 | 26 | 0.34 |
| GO:0000379 | tRNA-type intron splice site recognition and cleavage | 0.004999 | 27 | 0.36 |

| | | | | |
|---|---|---|---|---|
| GO:0001775 | cell activation | 0.005234 | 28 | 0.37 |
| GO:0051012 | microtubule sliding | 0.005393 | 29 | 0.38 |
| GO:1902856 | negative regulation of nonmotile primary cilium assembly | 0.005393 | 30 | 0.39 |
| GO:1902572 | negative regulation of serine-type peptidase activity | 0.006425 | 31 | 0.41 |
| GO:0045785 | positive regulation of cell adhesion | 0.007128 | 32 | 0.42 |
| GO:0030091 | protein repair | 0.007384 | 33 | 0.43 |
| GO:0014717 | regulation of satellite cell activation involved in skeletal muscle regeneration | 0.008105 | 34 | 0.45 |
| GO:1901899 | positive regulation of relaxation of cardiac muscle | 0.008212 | 35 | 0.46 |
| GO:1902571 | regulation of serine-type peptidase activity | 0.008889 | 36 | 0.47 |
| GO:0007267 | cell-cell signaling | 0.009077 | 37 | 0.49 |
| GO:0045217 | cell-cell junction maintenance | 0.010775 | 38 | 0.5 |
| GO:0010817 | regulation of hormone levels | 0.01084 | 39 | 0.51 |
| GO:0021853 | cerebral cortex GABAergic interneuron migration | 0.012358 | 40 | 0.53 |
| GO:0000394 | RNA splicing, via endonucleolytic cleavage and ligation | 0.013606 | 41 | 0.54 |
| GO:1903361 | protein localization to basolateral plasma membrane | 0.014024 | 42 | 0.55 |
| GO:0032609 | interferon-gamma production | 0.014738 | 43 | 0.57 |
| GO:0002505 | antigen processing and presentation of polysaccharide antigen via MHC class II | 0.01674 | 44 | 0.58 |
| GO:0002506 | polysaccharide assembly with MHC class II protein complex | 0.01674 | 45 | 0.59 |
| GO:0071823 | protein-carbohydrate complex subunit organization | 0.01674 | 46 | 0.61 |
| GO:0097021 | lymphocyte migration into lymphoid organs | 0.01725 | 47 | 0.62 |
| GO:0014719 | skeletal muscle satellite cell activation | 0.020694 | 48 | 0.63 |
| GO:0051262 | protein tetramerization | 0.022051 | 49 | 0.64 |
| GO:0071322 | cellular response to carbohydrate stimulus | 0.023286 | 50 | 0.66 |
| GO:0070208 | protein heterotrimerization | 0.02422 | 51 | 0.67 |
| GO:0043586 | tongue development | 0.024362 | 52 | 0.68 |
| GO:0061146 | Peyer's patch morphogenesis | 0.024482 | 53 | 0.7 |
| GO:0003416 | endochondral bone growth | 0.024653 | 54 | 0.71 |
| GO:0060046 | regulation of acrosome reaction | 0.025764 | 55 | 0.72 |

| | | | | |
|---|---|---|---|---|
| GO:0097154 | GABAergic neuron differentiation | 0.026122 | 56 | 0.74 |
| GO:0002396 | MHC protein complex assembly | 0.027389 | 57 | 0.75 |
| GO:0050663 | cytokine secretion | 0.029522 | 58 | 0.76 |
| GO:0007155 | cell adhesion | 0.030223 | 59 | 0.78 |
| GO:0048025 | negative regulation of mRNA splicing, via spliceosome | 0.030285 | 60 | 0.79 |
| GO:0050930 | induction of positive chemotaxis | 0.030321 | 61 | 0.8 |
| GO:0051641 | cellular localization | 0.030578 | 62 | 0.82 |
| GO:0022610 | biological adhesion | 0.030893 | 63 | 0.83 |
| GO:0006688 | glycosphingolipid biosynthetic process | 0.031241 | 64 | 0.84 |
| GO:0006887 | exocytosis | 0.031749 | 65 | 0.86 |
| GO:0060716 | labyrinthine layer blood vessel development | 0.032733 | 66 | 0.87 |
| GO:0050482 | arachidonic acid secretion | 0.033411 | 67 | 0.88 |
| GO:0043954 | cellular component maintenance | 0.036598 | 68 | 0.89 |
| GO:0010649 | regulation of cell communication by electrical coupling | 0.037733 | 69 | 0.91 |
| GO:0032844 | regulation of homeostatic process | 0.041279 | 70 | 0.92 |
| GO:0048537 | mucosal-associated lymphoid tissue development | 0.041796 | 71 | 0.93 |
| GO:0097503 | sialylation | 0.044311 | 72 | 0.95 |
| GO:0032846 | positive regulation of homeostatic process | 0.045836 | 73 | 0.96 |
| GO:0010882 | regulation of cardiac muscle contraction by calcium ion signaling | 0.047684 | 74 | 0.97 |
| GO:1902116 | negative regulation of organelle assembly | 0.048363 | 75 | 0.99 |
| GO:0006555 | methionine metabolic process | 0.049408 | 76 | 1 |

**Supplementary Data II-5: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in fetal brain**. P.DE refers to p-value for enrichment as computed by gometh() function in MissMethyl R package. Scaled Rank is computed as rank divided by the total number of terms in the list (n = 47).

| GO ID | Term | P.DE | Rank | ScaledRank |
|---|---|---|---|---|
| GO:0002480 | antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent | 0.000194 | 1 | 0.02 |
| GO:0009595 | detection of biotic stimulus | 0.000284 | 2 | 0.04 |
| GO:0016045 | detection of bacterium | 0.000284 | 3 | 0.06 |
| GO:0071346 | cellular response to interferon-gamma | 0.000338 | 4 | 0.09 |
| GO:0034341 | response to interferon-gamma | 0.000473 | 5 | 0.11 |
| GO:0019884 | antigen processing and presentation of exogenous antigen | 0.000658 | 6 | 0.13 |
| GO:0019882 | antigen processing and presentation | 0.001485 | 7 | 0.15 |
| GO:0032649 | regulation of interferon-gamma production | 0.001906 | 8 | 0.17 |
| GO:0032609 | interferon-gamma production | 0.002266 | 9 | 0.19 |
| GO:0035709 | memory T cell activation | 0.003379 | 10 | 0.21 |
| GO:2000568 | positive regulation of memory T cell activation | 0.003379 | 11 | 0.23 |
| GO:0019236 | response to pheromone | 0.003461 | 12 | 0.26 |
| GO:0002440 | production of molecular mediator of immune response | 0.004663 | 13 | 0.28 |
| GO:1901899 | positive regulation of relaxation of cardiac muscle | 0.005371 | 14 | 0.3 |
| GO:0043586 | tongue development | 0.007629 | 15 | 0.32 |
| GO:0042088 | T-helper 1 type immune response | 0.00818 | 16 | 0.34 |
| GO:1901844 | regulation of cell communication by electrical coupling involved in cardiac conduction | 0.009813 | 17 | 0.36 |
| GO:0002486 | antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent | 0.010066 | 18 | 0.38 |
| GO:0006388 | tRNA splicing, via endonucleolytic cleavage and ligation | 0.011328 | 19 | 0.4 |
| GO:2001179 | regulation of interleukin-10 secretion | 0.011902 | 20 | 0.43 |

| | | | | |
|---|---|---|---|---|
| GO:0042270 | protection from natural killer cell mediated cytotoxicity | 0.01277 | 21 | 0.45 |
| GO:0051481 | negative regulation of cytosolic calcium ion concentration | 0.013208 | 22 | 0.47 |
| GO:0048172 | regulation of short-term neuronal synaptic plasticity | 0.015344 | 23 | 0.49 |
| GO:0002369 | T cell cytokine production | 0.016892 | 24 | 0.51 |
| GO:0003416 | endochondral bone growth | 0.017508 | 25 | 0.53 |
| GO:0060314 | regulation of ryanodine-sensitive calcium-release channel activity | 0.019026 | 26 | 0.55 |
| GO:0022409 | positive regulation of cell-cell adhesion | 0.020701 | 27 | 0.57 |
| GO:0032633 | interleukin-4 production | 0.021045 | 28 | 0.6 |
| GO:0032673 | regulation of interleukin-4 production | 0.021045 | 29 | 0.62 |
| GO:0031342 | negative regulation of cell killing | 0.021288 | 30 | 0.64 |
| GO:0010644 | cell communication by electrical coupling | 0.021313 | 31 | 0.66 |
| GO:0035774 | positive regulation of insulin secretion involved in cellular response to glucose stimulus | 0.023315 | 32 | 0.68 |
| GO:0070296 | sarcoplasmic reticulum calcium ion transport | 0.023996 | 33 | 0.7 |
| GO:0008380 | RNA splicing | 0.024102 | 34 | 0.72 |
| GO:0032613 | interleukin-10 production | 0.026464 | 35 | 0.74 |
| GO:0001825 | blastocyst formation | 0.029925 | 36 | 0.77 |
| GO:0019731 | antibacterial humoral response | 0.030731 | 37 | 0.79 |
| GO:0002437 | inflammatory response to antigenic stimulus | 0.031129 | 38 | 0.81 |
| GO:0019883 | antigen processing and presentation of endogenous antigen | 0.033044 | 39 | 0.83 |
| GO:0019730 | antimicrobial humoral response | 0.033445 | 40 | 0.85 |
| GO:0051606 | detection of stimulus | 0.035732 | 41 | 0.87 |
| GO:0002704 | negative regulation of leukocyte mediated immunity | 0.03729 | 42 | 0.89 |
| GO:0006397 | mRNA processing | 0.037962 | 43 | 0.91 |
| GO:0002228 | natural killer cell mediated immunity | 0.039409 | 44 | 0.94 |
| GO:0009617 | response to bacterium | 0.046842 | 45 | 0.96 |
| GO:0071593 | lymphocyte aggregation | 0.04887 | 46 | 0.98 |
| GO:0060021 | palate development | 0.049548 | 47 | 1 |

**Supplementary Data II-6: meQTL evidence for every ASD-associated (PGC p-value < 1E-4) locus.**

Note: This table is too large to be included in this dissertation but can be found at https://www.nature.com/articles/s41467-017-00868-y

**Supplementary Data II-7: Enrichment Statistics comparing meQTL targets of cross-disorder PGC SNPs to meQTL targets of non cross disorder PGC associated SNPs**. Each cell represents the results of a Fisher's exact test examining the tendency of meQTL targets of significant (p <= 1E-4) SNPs (or their proxies) from the PGC cross disorder mega-analysis to overlap various functional regions compared to CpG sites that were also meQTL targets but not of signficant PGC cross disorder SNPs or their proxies. Green chromatin marks denote those typically associated with transcriptionally active chromatin; red marks denote those typically associated with inactive chromatin. We report the enrichment fold statistic or odds ratio and p-value. Values in parentheses in column headers denote the number of meQTL targets of significant PGC cross-disorder SNPs or their proxies in that tissue or that overlap in those two tissues (at FDR = 5% p-value for blood, cord blood, and lung and p-value = 1E-8 for fetal brain).

Note: This table is too large to be included in this dissertation but can be found at https://www.nature.com/articles/s41467-017-00868-y

**Supplementary Data II-8: Enrichment Statistics comparing meQTL targets to non-meQTL targets**. Each cell represents the results of a Fisher's exact test examining the tendency of meQTL targets in several groups to overlap various functional regions compared to non-meQTL targets. Green chromatin marks denote those typically associated with transcriptionally active chromatin; red marks denote those typically associated with inactive chromatin. We report the enrichment fold statistic or odds ratio and p-value. Values in parentheses in column headers denote the number of meQTL targets in that tissue or that overlap in those two tissues (at FDR = 5% p-value for blood, cord blood, and lung and p-value = 1E-8 for fetal brain).

Note: This table is too large to be included in this dissertation but can be found at https://www.nature.com/articles/s41467-017-00868-y

# APPENDIX III

**Table III-1: Summary of meQTL evidence across tissue type**. The number of SNPs and independent sites that fall under each category of cross-tissue meQTL presence. Results are shown for meQTLs associated at FDR = 1% threshold in blood, cord blood, and placenta datasets and threshold of 1E-8 in fetal brain. Only SNPs that were included in all three tissues in their respective meQTL queries are included in this analysis (n = 2,620,313). Independent sites were constructed by grouping SNPs into bins defined by recombination hot spot data from 1000 Genomes (see Methods). For example, scenario 1 lists that there are a total of 111,574 SNPs that are meQTLs in blood, cord blood, placenta, and fetal brain, which fall into 5,809 loci.

| Scenario | Blood | Cord Blood | Placenta | Fetal Brain | SNPs | % of Total SNPs | Independent Sites | % of Total Independent Sites |
|---|---|---|---|---|---|---|---|---|
| ***1*** | ✓ | ✓ | ✓ | ✓ | 111,574 | 4.26% | 5,809 | 3.70% |
| ***2*** | ✓ | ✓ | ✓ | × | 289,240 | 11.04% | 14,744 | 9.40% |
| ***3*** | ✓ | × | ✓ | ✓ | 32,409 | 1.24% | 1,381 | 0.88% |
| ***4*** | ✓ | ✓ | × | ✓ | 5,067 | 0.19% | 397 | 0.25% |
| ***5*** | ✓ | ✓ | × | × | 37,567 | 1.43% | 2,659 | 1.70% |
| ***6*** | ✓ | × | ✓ | × | 483,903 | 18.47% | 25,747 | 16.42% |
| ***7*** | ✓ | × | × | ✓ | 3,243 | 0.12% | 201 | 0.13% |
| ***8*** | ✓ | × | × | × | 129,549 | 4.94% | 9,613 | 6.13% |
| ***9*** | × | ✓ | ✓ | ✓ | 494 | 0.02% | 26 | 0.02% |
| ***10*** | × | ✓ | ✓ | × | 4,130 | 0.16% | 201 | 0.13% |
| ***11*** | × | ✓ | × | ✓ | 62 | 0.00% | 3 | 0.00% |
| ***12*** | × | ✓ | × | × | 1,059 | 0.04% | 73 | 0.05% |
| ***13*** | × | × | ✓ | ✓ | 5,959 | 0.23% | 274 | 0.17% |
| ***14*** | × | × | ✓ | × | 902,884 | 34.46% | 48,387 | 30.85% |
| ***15*** | × | × | × | ✓ | 1,418 | 0.05% | 95 | 0.06% |
| ***16*** | × | × | × | × | 611,755 | 23.35% | 47,230 | 30.11% |
| | | | | **SUM** | 2,620,313 | | 156,840 | |

**Table III-2: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in fetal brain.** P.DE refers to p-value for enrichment as computed by *gometh()* function in *MissMethyl* R package.

| GO ID | Term | P.DE |
|---|---|---|
| GO:0000459 | exonucleolytic trimming involved in rRNA processing | 0.001391 |
| GO:0000467 | exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) | 0.001391 |
| GO:1990000 | amyloid fibril formation | 0.001479 |
| GO:0048143 | astrocyte activation | 0.001479 |
| GO:1903748 | negative regulation of establishment of protein localization to mitochondrion | 0.001479 |
| GO:0090258 | negative regulation of mitochondrial fission | 0.001479 |
| GO:1902988 | neurofibrillary tangle assembly | 0.001479 |
| GO:1905689 | positive regulation of diacylglycerol kinase activity | 0.001479 |
| GO:1902474 | positive regulation of protein localization to synapse | 0.001479 |
| GO:1905687 | regulation of diacylglycerol kinase activity | 0.001479 |
| GO:1902473 | regulation of protein localization to synapse | 0.001479 |
| GO:0019518 | L-threonine catabolic process to glycine | 0.002113 |
| GO:0006567 | threonine catabolic process | 0.002113 |
| GO:0031125 | rRNA 3'-end processing | 0.002559 |
| GO:0048312 | intracellular distribution of mitochondria | 0.003157 |
| GO:0072386 | plus-end-directed organelle transport along microtubule | 0.003259 |
| GO:0032930 | positive regulation of superoxide anion generation | 0.003557 |
| GO:0006566 | threonine metabolic process | 0.00389 |
| GO:0048311 | mitochondrion distribution | 0.004695 |
| GO:0010917 | negative regulation of mitochondrial membrane potential | 0.004979 |
| GO:0019896 | axonal transport of mitochondrion | 0.005077 |
| GO:0043628 | ncRNA 3'-end processing | 0.005654 |
| GO:0045837 | negative regulation of membrane potential | 0.005889 |
| GO:0032928 | regulation of superoxide anion generation | 0.00625 |
| GO:0001774 | microglial cell activation | 0.006514 |
| GO:0006544 | glycine metabolic process | 0.006532 |
| GO:0051654 | establishment of mitochondrion localization | 0.008162 |
| GO:0034643 | establishment of mitochondrion localization, microtubule-mediated | 0.008162 |
| GO:0047497 | mitochondrion transport along microtubule | 0.008162 |
| GO:0007194 | negative regulation of adenylate cyclase activity | 0.008186 |
| GO:0090140 | regulation of mitochondrial fission | 0.008634 |
| GO:0090503 | RNA phosphodiester bond hydrolysis, exonucleolytic | 0.008766 |
| GO:0009068 | aspartate family amino acid catabolic process | 0.009062 |

| | | |
|---|---|---|
| GO:0031116 | positive regulation of microtubule polymerization | 0.009138 |
| GO:0031112 | positive regulation of microtubule polymerization or depolymerization | 0.009138 |
| GO:0031280 | negative regulation of cyclase activity | 0.009578 |
| GO:0014002 | astrocyte development | 0.009962 |
| GO:0000469 | cleavage involved in rRNA processing | 0.010114 |
| GO:0045936 | negative regulation of phosphate metabolic process | 0.010192 |
| GO:0010563 | negative regulation of phosphorus metabolic process | 0.010192 |
| GO:0051350 | negative regulation of lyase activity | 0.010646 |
| GO:0000466 | maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) | 0.010703 |
| GO:0045773 | positive regulation of axon extension | 0.010922 |
| GO:0090322 | regulation of superoxide metabolic process | 0.011023 |
| GO:0035418 | protein localization to synapse | 0.011712 |
| GO:0048708 | astrocyte differentiation | 0.011882 |
| GO:0042554 | superoxide anion generation | 0.012275 |
| GO:0051646 | mitochondrion localization | 0.012275 |
| GO:0042116 | macrophage activation | 0.012587 |
| GO:0030818 | negative regulation of cAMP biosynthetic process | 0.012637 |
| GO:0070841 | inclusion body assembly | 0.012642 |
| GO:0007214 | gamma-aminobutyric acid signaling pathway | 0.012721 |
| GO:0031113 | regulation of microtubule polymerization | 0.013371 |
| GO:0030815 | negative regulation of cAMP metabolic process | 0.013622 |
| GO:0031122 | cytoplasmic microtubule organization | 0.013756 |
| GO:0000266 | mitochondrial fission | 0.013822 |
| GO:0050853 | B cell receptor signaling pathway | 0.013887 |
| GO:0030803 | negative regulation of cyclic nucleotide biosynthetic process | 0.014024 |
| GO:0030809 | negative regulation of nucleotide biosynthetic process | 0.014024 |
| GO:1900372 | negative regulation of purine nucleotide biosynthetic process | 0.014024 |
| GO:0000460 | maturation of 5.8S rRNA | 0.014332 |
| GO:0009069 | serine family amino acid metabolic process | 0.014558 |
| GO:0030800 | negative regulation of cyclic nucleotide metabolic process | 0.015008 |
| GO:0010823 | negative regulation of mitochondrion organization | 0.016394 |
| GO:0098930 | axonal transport | 0.016464 |
| GO:0050772 | positive regulation of axonogenesis | 0.01682 |
| GO:0031110 | regulation of microtubule polymerization or depolymerization | 0.017216 |
| GO:0008088 | axo-dendritic transport | 0.017388 |
| GO:0046785 | microtubule polymerization | 0.01814 |
| GO:0043086 | negative regulation of catalytic activity | 0.019029 |
| GO:0072384 | organelle transport along microtubule | 0.019267 |

| | | |
|---|---|---|
| GO:0050848 | regulation of calcium-mediated signaling | 0.01963 |
| GO:0070201 | regulation of establishment of protein localization | 0.021081 |
| GO:0006801 | superoxide metabolic process | 0.021298 |
| GO:0009066 | aspartate family amino acid metabolic process | 0.021939 |
| GO:0032024 | positive regulation of insulin secretion | 0.022385 |
| GO:0009892 | negative regulation of metabolic process | 0.023193 |
| GO:0090277 | positive regulation of peptide hormone secretion | 0.023352 |
| GO:1900543 | negative regulation of purine nucleotide metabolic process | 0.023682 |
| GO:0007193 | adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway | 0.024375 |
| GO:0045980 | negative regulation of nucleotide metabolic process | 0.024684 |
| GO:0030516 | regulation of axon extension | 0.02492 |
| GO:0031109 | microtubule polymerization or depolymerization | 0.025531 |
| GO:0051881 | regulation of mitochondrial membrane potential | 0.026349 |
| GO:0038083 | peptidyl-tyrosine autophosphorylation | 0.026561 |
| GO:0021782 | glial cell development | 0.028007 |
| GO:0032880 | regulation of protein localization | 0.028023 |
| GO:0061387 | regulation of extent of cell growth | 0.028897 |
| GO:0048675 | axon extension | 0.029531 |
| GO:0045761 | regulation of adenylate cyclase activity | 0.029783 |
| GO:0046887 | positive regulation of hormone secretion | 0.029899 |
| GO:1903828 | negative regulation of cellular protein localization | 0.030214 |
| GO:0007613 | memory | 0.031193 |
| GO:0099111 | microtubule-based transport | 0.032655 |
| GO:0010970 | transport along microtubule | 0.032655 |
| GO:1903747 | regulation of establishment of protein localization to mitochondrion | 0.033539 |
| GO:0010770 | positive regulation of cell morphogenesis involved in differentiation | 0.033557 |
| GO:1901606 | alpha-amino acid catabolic process | 0.034177 |
| GO:0044092 | negative regulation of molecular function | 0.034199 |
| GO:1901216 | positive regulation of neuron death | 0.03447 |
| GO:0006919 | activation of cysteine-type endopeptidase activity involved in apoptotic process | 0.035086 |
| GO:2000379 | positive regulation of reactive oxygen species metabolic process | 0.035188 |
| GO:1900034 | regulation of cellular response to heat | 0.035714 |
| GO:0031279 | regulation of cyclase activity | 0.035981 |
| GO:0006796 | phosphate-containing compound metabolic process | 0.035984 |
| GO:0051339 | regulation of lyase activity | 0.036125 |
| GO:0009063 | cellular amino acid catabolic process | 0.036205 |
| GO:0006171 | cAMP biosynthetic process | 0.036764 |

| | | |
|---|---|---|
| GO:0030817 | regulation of cAMP biosynthetic process | 0.036764 |
| GO:0006793 | phosphorus metabolic process | 0.03795 |
| GO:0030705 | cytoskeleton-dependent intracellular transport | 0.038996 |
| GO:0031047 | gene silencing by RNA | 0.039264 |
| GO:0031123 | RNA 3'-end processing | 0.039891 |
| GO:0009190 | cyclic nucleotide biosynthetic process | 0.040056 |
| GO:0052652 | cyclic purine nucleotide metabolic process | 0.040056 |
| GO:0030802 | regulation of cyclic nucleotide biosynthetic process | 0.040056 |
| GO:0030808 | regulation of nucleotide biosynthetic process | 0.040056 |
| GO:1900371 | regulation of purine nucleotide biosynthetic process | 0.040056 |
| GO:0030814 | regulation of cAMP metabolic process | 0.04146 |
| GO:0019722 | calcium-mediated signaling | 0.041882 |
| GO:0070507 | regulation of microtubule cytoskeleton organization | 0.042236 |
| GO:2001020 | regulation of response to DNA damage stimulus | 0.04225 |
| GO:0050770 | regulation of axonogenesis | 0.042585 |
| GO:0032273 | positive regulation of protein polymerization | 0.04261 |
| GO:0043280 | positive regulation of cysteine-type endopeptidase activity involved in apoptotic process | 0.044364 |
| GO:0030799 | regulation of cyclic nucleotide metabolic process | 0.044738 |
| GO:0072655 | establishment of protein localization to mitochondrion | 0.045406 |
| GO:2001056 | positive regulation of cysteine-type endopeptidase activity | 0.045423 |
| GO:1990138 | neuron projection extension | 0.045583 |
| GO:0010001 | glial cell differentiation | 0.046054 |
| GO:0046058 | cAMP metabolic process | 0.046621 |
| GO:0007188 | adenylate cyclase-modulating G-protein coupled receptor signaling pathway | 0.047007 |
| GO:0070585 | protein localization to mitochondrion | 0.047026 |
| GO:0030307 | positive regulation of cell growth | 0.047315 |
| GO:0010950 | positive regulation of endopeptidase activity | 0.047851 |
| GO:0048639 | positive regulation of developmental growth | 0.049279 |
| GO:0032886 | regulation of microtubule-based process | 0.049754 |

**Table III-3: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in peripheral blood.** P.DE refers to p-value for enrichment as computed by *gometh()* function in *MissMethyl* R package.

| GO ID | Term | P.DE |
|---|---|---|
| GO:0010976 | positive regulation of neuron projection development | 0.002244 |
| GO:1903944 | negative regulation of hepatocyte apoptotic process | 0.002667 |
| GO:1901740 | negative regulation of myoblast fusion | 0.002667 |
| GO:1903943 | regulation of hepatocyte apoptotic process | 0.002667 |
| GO:1905474 | canonical Wnt signaling pathway involved in stem cell proliferation | 0.00345 |
| GO:0060064 | Spemann organizer formation at the anterior end of the primitive streak | 0.00345 |
| GO:1904744 | positive regulation of telomeric DNA binding | 0.003606 |
| GO:0030307 | positive regulation of cell growth | 0.003877 |
| GO:1990138 | neuron projection extension | 0.004527 |
| GO:0045338 | farnesyl diphosphate metabolic process | 0.004586 |
| GO:0044108 | cellular alcohol biosynthetic process | 0.004586 |
| GO:0044107 | cellular alcohol metabolic process | 0.004586 |
| GO:0006696 | ergosterol biosynthetic process | 0.004586 |
| GO:0008204 | ergosterol metabolic process | 0.004586 |
| GO:0016129 | phytosteroid biosynthetic process | 0.004586 |
| GO:0016128 | phytosteroid metabolic process | 0.004586 |
| GO:0032413 | negative regulation of ion transmembrane transporter activity | 0.005059 |
| GO:0035934 | corticosterone secretion | 0.005255 |
| GO:2000852 | regulation of corticosterone secretion | 0.005255 |
| GO:0048639 | positive regulation of developmental growth | 0.005342 |
| GO:1904063 | negative regulation of cation transmembrane transport | 0.005813 |
| GO:0019518 | L-threonine catabolic process to glycine | 0.005905 |
| GO:0006567 | threonine catabolic process | 0.005905 |
| GO:0034242 | negative regulation of syncytium formation by plasma membrane fusion | 0.006124 |
| GO:0045666 | positive regulation of neuron differentiation | 0.006242 |
| GO:1905689 | positive regulation of diacylglycerol kinase activity | 0.006458 |
| GO:1905687 | regulation of diacylglycerol kinase activity | 0.006458 |
| GO:0009066 | aspartate family amino acid metabolic process | 0.006466 |

| | | |
|---|---|---|
| GO:0071376 | cellular response to corticotropin-releasing hormone stimulus | 0.006867 |
| GO:0043435 | response to corticotropin-releasing hormone | 0.006867 |
| GO:0060061 | Spemann organizer formation | 0.006881 |
| GO:0031346 | positive regulation of cell projection organization | 0.006986 |
| GO:2001288 | positive regulation of caveolin-mediated endocytosis | 0.007227 |
| GO:1904751 | positive regulation of protein localization to nucleolus | 0.007547 |
| GO:0048697 | positive regulation of collateral sprouting in absence of injury | 0.007722 |
| GO:0048696 | regulation of collateral sprouting in absence of injury | 0.007722 |
| GO:0032410 | negative regulation of transporter activity | 0.007849 |
| GO:0051458 | corticotropin secretion | 0.007862 |
| GO:1902988 | neurofibrillary tangle assembly | 0.008241 |
| GO:0045008 | depyrimidination | 0.008417 |
| GO:0006285 | base-excision repair, AP site formation | 0.008742 |
| GO:0034766 | negative regulation of ion transmembrane transport | 0.009037 |
| GO:0044338 | canonical Wnt signaling pathway involved in mesenchymal stem cell differentiation | 0.009238 |
| GO:0006566 | threonine metabolic process | 0.009529 |
| GO:0000459 | exonucleolytic trimming involved in rRNA processing | 0.009643 |
| GO:0000467 | exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) | 0.009643 |
| GO:1904742 | regulation of telomeric DNA binding | 0.009767 |
| GO:1990000 | amyloid fibril formation | 0.010146 |
| GO:1904954 | canonical Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation | 0.010299 |
| GO:0048588 | developmental cell growth | 0.010354 |
| GO:0034763 | negative regulation of transmembrane transport | 0.010617 |
| GO:0048143 | astrocyte activation | 0.010981 |
| GO:0031125 | rRNA 3'-end processing | 0.011038 |
| GO:1904749 | regulation of protein localization to nucleolus | 0.011184 |
| GO:0097384 | cellular lipid biosynthetic process | 0.011208 |
| GO:0050769 | positive regulation of neurogenesis | 0.011296 |

| | | |
|---|---|---|
| GO:0072126 | positive regulation of glomerular mesangial cell proliferation | 0.011449 |
| GO:1902570 | protein localization to nucleolus | 0.012034 |
| GO:0048312 | intracellular distribution of mitochondria | 0.012359 |
| GO:0044339 | canonical Wnt signaling pathway involved in osteoblast differentiation | 0.012709 |
| GO:0060560 | developmental growth involved in morphogenesis | 0.013214 |
| GO:0050772 | positive regulation of axonogenesis | 0.013303 |
| GO:0014732 | skeletal muscle atrophy | 0.013363 |
| GO:0014891 | striated muscle atrophy | 0.013363 |
| GO:0045927 | positive regulation of growth | 0.01396 |
| GO:2000347 | positive regulation of hepatocyte proliferation | 0.014205 |
| GO:0090258 | negative regulation of mitochondrial fission | 0.014302 |
| GO:2001286 | regulation of caveolin-mediated endocytosis | 0.014647 |
| GO:0030516 | regulation of axon extension | 0.014911 |
| GO:1903748 | negative regulation of establishment of protein localization to mitochondrion | 0.015266 |
| GO:0014842 | regulation of skeletal muscle satellite cell proliferation | 0.015531 |
| GO:0010975 | regulation of neuron projection development | 0.01572 |
| GO:0009223 | pyrimidine deoxyribonucleotide catabolic process | 0.015729 |
| GO:0045761 | regulation of adenylate cyclase activity | 0.016017 |
| GO:0048669 | collateral sprouting in absence of injury | 0.016873 |
| GO:0030091 | protein repair | 0.01693 |
| GO:0090193 | positive regulation of glomerulus development | 0.01712 |
| GO:1901724 | positive regulation of cell proliferation involved in kidney development | 0.017182 |
| GO:0044340 | canonical Wnt signaling pathway involved in regulation of cell proliferation | 0.017339 |
| GO:0048311 | mitochondrion distribution | 0.017531 |
| GO:0014866 | skeletal myofibril assembly | 0.017545 |
| GO:0010720 | positive regulation of cell development | 0.017596 |
| GO:0043628 | ncRNA 3'-end processing | 0.018464 |
| GO:0072386 | plus-end-directed organelle transport along microtubule | 0.018529 |
| GO:0006244 | pyrimidine nucleotide catabolic process | 0.018637 |
| GO:0014841 | skeletal muscle satellite cell proliferation | 0.018905 |
| GO:1901386 | negative regulation of voltage-gated calcium channel activity | 0.019063 |

| | | |
|---|---|---|
| GO:0051962 | positive regulation of nervous system development | 0.019615 |
| GO:0031279 | regulation of cyclase activity | 0.020282 |
| GO:0061387 | regulation of extent of cell growth | 0.020375 |
| GO:0014857 | regulation of skeletal muscle cell proliferation | 0.020672 |
| GO:0051339 | regulation of lyase activity | 0.020767 |
| GO:2000009 | negative regulation of protein localization to cell surface | 0.020881 |
| GO:1902306 | negative regulation of sodium ion transmembrane transport | 0.02129 |
| GO:2000650 | negative regulation of sodium ion transmembrane transporter activity | 0.02129 |
| GO:0035933 | glucocorticoid secretion | 0.021527 |
| GO:2000849 | regulation of glucocorticoid secretion | 0.021527 |
| GO:0032211 | negative regulation of telomere maintenance via telomerase | 0.021735 |
| GO:0072124 | regulation of glomerular mesangial cell proliferation | 0.021807 |
| GO:0071732 | cellular response to nitric oxide | 0.022146 |
| GO:0019896 | axonal transport of mitochondrion | 0.022424 |
| GO:0043981 | histone H4-K5 acetylation | 0.022479 |
| GO:0043982 | histone H4-K8 acetylation | 0.022479 |
| GO:0048666 | neuron development | 0.022945 |
| GO:0048675 | axon extension | 0.023532 |
| GO:0010579 | positive regulation of adenylate cyclase activity involved in G-protein coupled receptor signaling pathway | 0.023714 |
| GO:0010578 | regulation of adenylate cyclase activity involved in G-protein coupled receptor signaling pathway | 0.023714 |
| GO:0072584 | caveolin-mediated endocytosis | 0.023727 |
| GO:0014856 | skeletal muscle cell proliferation | 0.024029 |
| GO:0006883 | cellular sodium ion homeostasis | 0.024351 |
| GO:0048638 | regulation of developmental growth | 0.024425 |
| GO:0010917 | negative regulation of mitochondrial membrane potential | 0.02462 |
| GO:0043271 | negative regulation of ion transport | 0.024625 |
| GO:0060174 | limb bud formation | 0.024744 |
| GO:0014889 | muscle atrophy | 0.025036 |
| GO:0010766 | negative regulation of sodium ion transport | 0.025336 |
| GO:0072110 | glomerular mesangial cell proliferation | 0.025376 |

| | | |
|---|---|---|
| GO:1902474 | positive regulation of protein localization to synapse | 0.02546 |
| GO:1902473 | regulation of protein localization to synapse | 0.02546 |
| GO:0097284 | hepatocyte apoptotic process | 0.025485 |
| GO:0071731 | response to nitric oxide | 0.025686 |
| GO:2000345 | regulation of hepatocyte proliferation | 0.025919 |
| GO:0032930 | positive regulation of superoxide anion generation | 0.026001 |
| GO:0009264 | deoxyribonucleotide catabolic process | 0.026022 |
| GO:0046386 | deoxyribose phosphate catabolic process | 0.026022 |
| GO:0048672 | positive regulation of collateral sprouting | 0.02606 |
| GO:0090192 | regulation of glomerulus development | 0.02742 |
| GO:0048843 | negative regulation of axon extension involved in axon guidance | 0.027571 |
| GO:0009219 | pyrimidine deoxyribonucleotide metabolic process | 0.027646 |
| GO:0043280 | positive regulation of cysteine-type endopeptidase activity involved in apoptotic process | 0.027707 |
| GO:0032928 | regulation of superoxide anion generation | 0.029548 |
| GO:0006544 | glycine metabolic process | 0.029966 |
| GO:0045837 | negative regulation of membrane potential | 0.030087 |
| GO:1902170 | cellular response to reactive nitrogen species | 0.030179 |
| GO:2001056 | positive regulation of cysteine-type endopeptidase activity | 0.030342 |
| GO:1902668 | negative regulation of axon guidance | 0.030382 |
| GO:0010832 | negative regulation of myotube differentiation | 0.030754 |
| GO:0043984 | histone H4-K16 acetylation | 0.030825 |
| GO:1901722 | regulation of cell proliferation involved in kidney development | 0.030939 |
| GO:1904953 | Wnt signaling pathway involved in midbrain dopaminergic neuron differentiation | 0.032022 |
| GO:0051974 | negative regulation of telomerase activity | 0.032037 |
| GO:0060544 | regulation of necroptotic process | 0.032064 |
| GO:0000466 | maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) | 0.03232 |
| GO:0001774 | microglial cell activation | 0.032607 |
| GO:0051129 | negative regulation of cellular component organization | 0.032883 |
| GO:0030817 | regulation of cAMP biosynthetic process | 0.032908 |

| | | |
|---|---|---|
| GO:0072575 | epithelial cell proliferation involved in liver morphogenesis | 0.03364 |
| GO:0072574 | hepatocyte proliferation | 0.03364 |
| GO:0006171 | cAMP biosynthetic process | 0.033713 |
| GO:1901739 | regulation of myoblast fusion | 0.033936 |
| GO:1903828 | negative regulation of cellular protein localization | 0.034803 |
| GO:1903427 | negative regulation of reactive oxygen species biosynthetic process | 0.035122 |
| GO:0048589 | developmental growth | 0.035222 |
| GO:0010950 | positive regulation of endopeptidase activity | 0.035271 |
| GO:1901017 | negative regulation of potassium ion transmembrane transporter activity | 0.0356 |
| GO:0035930 | corticosteroid hormone secretion | 0.035966 |
| GO:2000846 | regulation of corticosteroid hormone secretion | 0.035966 |
| GO:0008299 | isoprenoid biosynthetic process | 0.036425 |
| GO:0007567 | parturition | 0.037001 |
| GO:0034643 | establishment of mitochondrion localization, microtubule-mediated | 0.037073 |
| GO:0047497 | mitochondrion transport along microtubule | 0.037073 |
| GO:0001704 | formation of primary germ layer | 0.037562 |
| GO:0009068 | aspartate family amino acid catabolic process | 0.037635 |
| GO:0072576 | liver morphogenesis | 0.037735 |
| GO:0072497 | mesenchymal stem cell differentiation | 0.038203 |
| GO:1901020 | negative regulation of calcium ion transmembrane transporter activity | 0.038871 |
| GO:0010667 | negative regulation of cardiac muscle cell apoptotic process | 0.039004 |
| GO:0072109 | glomerular mesangium development | 0.03921 |
| GO:1904357 | negative regulation of telomere maintenance via telomere lengthening | 0.039571 |
| GO:0051654 | establishment of mitochondrion localization | 0.039936 |
| GO:0030802 | regulation of cyclic nucleotide biosynthetic process | 0.039972 |
| GO:0031344 | regulation of cell projection organization | 0.040585 |
| GO:0043501 | skeletal muscle adaptation | 0.040741 |
| GO:0010770 | positive regulation of cell morphogenesis involved in differentiation | 0.041136 |
| GO:0030814 | regulation of cAMP metabolic process | 0.041718 |
| GO:1903170 | negative regulation of calcium ion transmembrane transport | 0.041965 |

| | | |
|---|---|---|
| GO:0000460 | maturation of 5.8S rRNA | 0.041969 |
| GO:0009950 | dorsal/ventral axis specification | 0.042 |
| GO:0009890 | negative regulation of biosynthetic process | 0.042137 |
| GO:0090503 | RNA phosphodiester bond hydrolysis, exonucleolytic | 0.042248 |
| GO:0010664 | negative regulation of striated muscle cell apoptotic process | 0.042319 |
| GO:0090322 | regulation of superoxide metabolic process | 0.042876 |
| GO:2000377 | regulation of reactive oxygen species metabolic process | 0.043127 |
| GO:0045664 | regulation of neuron differentiation | 0.043179 |
| GO:0010952 | positive regulation of peptidase activity | 0.043389 |
| GO:0000469 | cleavage involved in rRNA processing | 0.04405 |
| GO:0009190 | cyclic nucleotide biosynthetic process | 0.044074 |
| GO:0052652 | cyclic purine nucleotide metabolic process | 0.044074 |
| GO:1901380 | negative regulation of potassium ion transmembrane transport | 0.044086 |
| GO:0070198 | protein localization to chromosome, telomeric region | 0.04426 |
| GO:2000831 | regulation of steroid hormone secretion | 0.044276 |
| GO:0035929 | steroid hormone secretion | 0.044276 |
| GO:0072111 | cell proliferation involved in kidney development | 0.045288 |
| GO:0031116 | positive regulation of microtubule polymerization | 0.045883 |
| GO:0008361 | regulation of cell size | 0.046025 |
| GO:1900371 | regulation of purine nucleotide biosynthetic process | 0.046116 |
| GO:0009394 | 2'-deoxyribonucleotide metabolic process | 0.04653 |
| GO:0019692 | deoxyribose phosphate metabolic process | 0.04653 |
| GO:0042554 | superoxide anion generation | 0.04658 |
| GO:0090140 | regulation of mitochondrial fission | 0.046653 |
| GO:0030808 | regulation of nucleotide biosynthetic process | 0.046927 |
| GO:0001558 | regulation of cell growth | 0.047192 |
| GO:1903861 | positive regulation of dendrite extension | 0.047624 |
| GO:0048841 | regulation of axon extension involved in axon guidance | 0.047822 |
| GO:0007188 | adenylate cyclase-modulating G-protein coupled receptor signaling pathway | 0.048497 |
| GO:0060142 | regulation of syncytium formation by plasma membrane fusion | 0.048719 |

| | | |
|---|---|---|
| GO:1903055 | positive regulation of extracellular matrix organization | 0.049434 |
| GO:0010939 | regulation of necrotic cell death | 0.049848 |

**Table III-4: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in cord blood.** P.DE refers to p-value for enrichment as computed by *gometh()* function in *MissMethyl* R package.

| GO ID | Term | P.DE |
|---|---|---|
| GO:0044108 | cellular alcohol biosynthetic process | 0.002662 |
| GO:0044107 | cellular alcohol metabolic process | 0.002662 |
| GO:0006696 | ergosterol biosynthetic process | 0.002662 |
| GO:0008204 | ergosterol metabolic process | 0.002662 |
| GO:0045338 | farnesyl diphosphate metabolic process | 0.002662 |
| GO:0016129 | phytosteroid biosynthetic process | 0.002662 |
| GO:0016128 | phytosteroid metabolic process | 0.002662 |
| GO:0043280 | positive regulation of cysteine-type endopeptidase activity involved in apoptotic process | 0.003229 |
| GO:2001056 | positive regulation of cysteine-type endopeptidase activity | 0.003486 |
| GO:1990138 | neuron projection extension | 0.003623 |
| GO:1902988 | neurofibrillary tangle assembly | 0.003711 |
| GO:1905689 | positive regulation of diacylglycerol kinase activity | 0.003711 |
| GO:1905687 | regulation of diacylglycerol kinase activity | 0.003711 |
| GO:0010950 | positive regulation of endopeptidase activity | 0.003917 |
| GO:2001288 | positive regulation of caveolin-mediated endocytosis | 0.004093 |
| GO:0097384 | cellular lipid biosynthetic process | 0.004192 |
| GO:0090258 | negative regulation of mitochondrial fission | 0.004346 |
| GO:2001286 | regulation of caveolin-mediated endocytosis | 0.004803 |
| GO:0051345 | positive regulation of hydrolase activity | 0.004888 |
| GO:0010952 | positive regulation of peptidase activity | 0.004905 |
| GO:0048639 | positive regulation of developmental growth | 0.005188 |
| GO:1990000 | amyloid fibril formation | 0.005378 |
| GO:0006883 | cellular sodium ion homeostasis | 0.005816 |
| GO:1903748 | negative regulation of establishment of protein localization to mitochondrion | 0.005872 |
| GO:0030307 | positive regulation of cell growth | 0.005999 |
| GO:0048143 | astrocyte activation | 0.006494 |
| GO:1903828 | negative regulation of cellular protein localization | 0.006535 |
| GO:0072386 | plus-end-directed organelle transport along microtubule | 0.007035 |
| GO:0043981 | histone H4-K5 acetylation | 0.00733 |
| GO:0043982 | histone H4-K8 acetylation | 0.00733 |
| GO:0043281 | regulation of cysteine-type endopeptidase activity involved in apoptotic process | 0.00772 |
| GO:0043984 | histone H4-K16 acetylation | 0.007729 |

| | | |
|---|---|---|
| GO:0048588 | developmental cell growth | 0.007756 |
| GO:0019896 | axonal transport of mitochondrion | 0.007856 |
| GO:1902474 | positive regulation of protein localization to synapse | 0.008461 |
| GO:1902473 | regulation of protein localization to synapse | 0.008461 |
| GO:0051654 | establishment of mitochondrion localization | 0.008538 |
| GO:0034643 | establishment of mitochondrion localization, microtubule-mediated | 0.008538 |
| GO:0047497 | mitochondrion transport along microtubule | 0.008538 |
| GO:0060560 | developmental growth involved in morphogenesis | 0.008583 |
| GO:2000116 | regulation of cysteine-type endopeptidase activity | 0.008616 |
| GO:2000009 | negative regulation of protein localization to cell surface | 0.009069 |
| GO:0065009 | regulation of molecular function | 0.009501 |
| GO:0048312 | intracellular distribution of mitochondria | 0.009524 |
| GO:1901017 | negative regulation of potassium ion transmembrane transporter activity | 0.010909 |
| GO:0010917 | negative regulation of mitochondrial membrane potential | 0.01134 |
| GO:0001774 | microglial cell activation | 0.011414 |
| GO:0045927 | positive regulation of growth | 0.011608 |
| GO:0048311 | mitochondrion distribution | 0.012047 |
| GO:1902306 | negative regulation of sodium ion transmembrane transport | 0.01231 |
| GO:2000650 | negative regulation of sodium ion transmembrane transporter activity | 0.01231 |
| GO:0010766 | negative regulation of sodium ion transport | 0.012836 |
| GO:0051336 | regulation of hydrolase activity | 0.012927 |
| GO:0072584 | caveolin-mediated endocytosis | 0.013271 |
| GO:0048638 | regulation of developmental growth | 0.014013 |
| GO:1903861 | positive regulation of dendrite extension | 0.014028 |
| GO:1903859 | regulation of dendrite extension | 0.014028 |
| GO:0010976 | positive regulation of neuron projection development | 0.014505 |
| GO:0045837 | negative regulation of membrane potential | 0.014517 |
| GO:0008299 | isoprenoid biosynthetic process | 0.01462 |
| GO:1901380 | negative regulation of potassium ion transmembrane transport | 0.015179 |
| GO:0090140 | regulation of mitochondrial fission | 0.015355 |
| GO:0032930 | positive regulation of superoxide anion generation | 0.015883 |
| GO:0032928 | regulation of superoxide anion generation | 0.016347 |
| GO:0050790 | regulation of catalytic activity | 0.016587 |
| GO:0097484 | dendrite extension | 0.016698 |
| GO:0098930 | axonal transport | 0.017406 |
| GO:0014002 | astrocyte development | 0.018291 |
| GO:0070841 | inclusion body assembly | 0.018339 |
| GO:0052548 | regulation of endopeptidase activity | 0.020495 |

| | | |
|---|---|---|
| GO:0051646 | mitochondrion localization | 0.0205 |
| GO:0043267 | negative regulation of potassium ion transport | 0.021921 |
| GO:0045862 | positive regulation of proteolysis | 0.022254 |
| GO:0043085 | positive regulation of catalytic activity | 0.022919 |
| GO:0045666 | positive regulation of neuron differentiation | 0.023435 |
| GO:0052547 | regulation of peptidase activity | 0.023626 |
| GO:0042554 | superoxide anion generation | 0.024071 |
| GO:0007194 | negative regulation of adenylate cyclase activity | 0.024392 |
| GO:0031280 | negative regulation of cyclase activity | 0.024392 |
| GO:0031116 | positive regulation of microtubule polymerization | 0.02496 |
| GO:0007214 | gamma-aminobutyric acid signaling pathway | 0.025053 |
| GO:0032467 | positive regulation of cytokinesis | 0.025339 |
| GO:2000008 | regulation of protein localization to cell surface | 0.02547 |
| GO:0008088 | axo-dendritic transport | 0.025526 |
| GO:0010823 | negative regulation of mitochondrion organization | 0.025551 |
| GO:0055078 | sodium ion homeostasis | 0.025642 |
| GO:0000266 | mitochondrial fission | 0.025833 |
| GO:0060306 | regulation of membrane repolarization | 0.026395 |
| GO:0090322 | regulation of superoxide metabolic process | 0.026451 |
| GO:0051350 | negative regulation of lyase activity | 0.027288 |
| GO:0050853 | B cell receptor signaling pathway | 0.027829 |
| GO:0086009 | membrane repolarization | 0.028288 |
| GO:0031346 | positive regulation of cell projection organization | 0.028524 |
| GO:0042116 | macrophage activation | 0.028561 |
| GO:0006695 | cholesterol biosynthetic process | 0.029215 |
| GO:1902653 | secondary alcohol biosynthetic process | 0.029215 |
| GO:0070936 | protein K48-linked ubiquitination | 0.029337 |
| GO:1901016 | regulation of potassium ion transmembrane transporter activity | 0.029455 |
| GO:0050769 | positive regulation of neurogenesis | 0.030632 |
| GO:0032880 | regulation of protein localization | 0.030752 |
| GO:0086005 | ventricular cardiac muscle cell action potential | 0.031326 |
| GO:0031112 | positive regulation of microtubule polymerization or depolymerization | 0.031797 |
| GO:0003254 | regulation of membrane depolarization | 0.03187 |
| GO:0045773 | positive regulation of axon extension | 0.032106 |
| GO:0044092 | negative regulation of molecular function | 0.03228 |
| GO:0006513 | protein monoubiquitination | 0.032499 |
| GO:0030818 | negative regulation of cAMP biosynthetic process | 0.0326 |
| GO:0030803 | negative regulation of cyclic nucleotide biosynthetic process | 0.0326 |
| GO:0001558 | regulation of cell growth | 0.03281 |

| | | |
|---|---|---|
| GO:0031113 | regulation of microtubule polymerization | 0.03442 |
| GO:0030809 | negative regulation of nucleotide biosynthetic process | 0.034587 |
| GO:1900372 | negative regulation of purine nucleotide biosynthetic process | 0.034587 |
| GO:0030815 | negative regulation of cAMP metabolic process | 0.034826 |
| GO:0030800 | negative regulation of cyclic nucleotide metabolic process | 0.034826 |
| GO:0072384 | organelle transport along microtubule | 0.035202 |
| GO:0016126 | sterol biosynthetic process | 0.036787 |
| GO:0043967 | histone H4 acetylation | 0.037702 |
| GO:0034394 | protein localization to cell surface | 0.037777 |
| GO:0035418 | protein localization to synapse | 0.038496 |
| GO:0044093 | positive regulation of molecular function | 0.03893 |
| GO:0001706 | endoderm formation | 0.038982 |
| GO:2000649 | regulation of sodium ion transmembrane transporter activity | 0.039899 |
| GO:0032413 | negative regulation of ion transmembrane transporter activity | 0.040227 |
| GO:0045934 | negative regulation of nucleobase-containing compound metabolic process | 0.041205 |
| GO:0010975 | regulation of neuron projection development | 0.041796 |
| GO:0042391 | regulation of membrane potential | 0.041935 |
| GO:0010720 | positive regulation of cell development | 0.042406 |
| GO:0022604 | regulation of cell morphogenesis | 0.042883 |
| GO:1904063 | negative regulation of cation transmembrane transport | 0.042933 |
| GO:0051962 | positive regulation of nervous system development | 0.043027 |
| GO:0048708 | astrocyte differentiation | 0.043568 |
| GO:0086002 | cardiac muscle cell action potential involved in contraction | 0.044143 |
| GO:0031122 | cytoplasmic microtubule organization | 0.044494 |
| GO:1902305 | regulation of sodium ion transmembrane transport | 0.044882 |
| GO:0051781 | positive regulation of cell division | 0.045409 |
| GO:0050772 | positive regulation of axonogenesis | 0.045825 |
| GO:0016049 | cell growth | 0.045833 |
| GO:0051881 | regulation of mitochondrial membrane potential | 0.046107 |
| GO:0046785 | microtubule polymerization | 0.046366 |
| GO:0030104 | water homeostasis | 0.046827 |
| GO:0050848 | regulation of calcium-mediated signaling | 0.047181 |
| GO:0006801 | superoxide metabolic process | 0.049141 |
| GO:1900034 | regulation of cellular response to heat | 0.049216 |

**Table III-5: Marginally significant Gene Ontology Terms post REVIGO comparing ASD-related meQTL targets to meQTL targets generally in placenta.** P.DE refers to p-value for enrichment as computed by *gometh()* function in *MissMethyl* R package.

| GO ID | Term | P.DE |
|---|---|---|
| GO:1904744 | positive regulation of telomeric DNA binding | 0.000407 |
| GO:0039506 | modulation by virus of host molecular function | 0.00054 |
| GO:0039507 | suppression by virus of host molecular function | 0.00054 |
| GO:0039513 | suppression by virus of host catalytic activity | 0.00054 |
| GO:0039516 | modulation by virus of host catalytic activity | 0.00054 |
| GO:0039650 | suppression by virus of host cysteine-type endopeptidase activity involved in apoptotic process | 0.00054 |
| GO:0052053 | negative regulation by symbiont of host catalytic activity | 0.00054 |
| GO:0052055 | modulation by symbiont of host molecular function | 0.00054 |
| GO:0052056 | negative regulation by symbiont of host molecular function | 0.00054 |
| GO:0052148 | modulation by symbiont of host catalytic activity | 0.00054 |
| GO:0052199 | negative regulation of catalytic activity in other organism involved in symbiotic interaction | 0.00054 |
| GO:0098939 | dendritic transport of mitochondrion | 0.000664 |
| GO:0098972 | anterograde dendritic transport of mitochondrion | 0.000664 |
| GO:0005979 | regulation of glycogen biosynthetic process | 0.001391 |
| GO:0010962 | regulation of glucan biosynthetic process | 0.001391 |
| GO:0019050 | suppression by virus of host apoptotic process | 0.00161 |
| GO:0039526 | modulation by virus of host apoptotic process | 0.00161 |
| GO:0032885 | regulation of polysaccharide biosynthetic process | 0.002221 |
| GO:0033668 | negative regulation by symbiont of host apoptotic process | 0.002259 |
| GO:0052041 | negative regulation by symbiont of host programmed cell death | 0.002259 |
| GO:0052490 | negative regulation by organism of programmed cell death in other organism involved in symbiotic interaction | 0.002259 |
| GO:0070873 | regulation of glycogen metabolic process | 0.002439 |
| GO:1904742 | regulation of telomeric DNA binding | 0.002984 |
| GO:0032881 | regulation of polysaccharide metabolic process | 0.003495 |
| GO:0052203 | modulation of catalytic activity in other organism involved in symbiotic interaction | 0.003685 |
| GO:0060715 | syncytiotrophoblast cell differentiation involved in labyrinthine layer development | 0.003696 |
| GO:0045494 | photoreceptor cell maintenance | 0.003762 |
| GO:0052150 | modulation by symbiont of host apoptotic process | 0.004051 |
| GO:0005978 | glycogen biosynthetic process | 0.00425 |
| GO:0009250 | glucan biosynthetic process | 0.00425 |

| | | |
|---|---|---|
| GO:0090503 | RNA phosphodiester bond hydrolysis, exonucleolytic | 0.004558 |
| GO:0006397 | mRNA processing | 0.004727 |
| GO:1903527 | positive regulation of membrane tubulation | 0.00522 |
| GO:0046677 | response to antibiotic | 0.005269 |
| GO:0098935 | dendritic transport | 0.005487 |
| GO:0098937 | anterograde dendritic transport | 0.005487 |
| GO:0090282 | positive regulation of transcription involved in G2/M transition of mitotic cell cycle | 0.005825 |
| GO:0006396 | RNA processing | 0.005891 |
| GO:2001288 | positive regulation of caveolin-mediated endocytosis | 0.005944 |
| GO:0000377 | RNA splicing, via transesterification reactions with bulged adenosine as nucleophile | 0.006155 |
| GO:0000398 | mRNA splicing, via spliceosome | 0.006155 |
| GO:0000375 | RNA splicing, via transesterification reactions | 0.006285 |
| GO:0071931 | positive regulation of transcription involved in G1/S transition of mitotic cell cycle | 0.006339 |
| GO:0044531 | modulation of programmed cell death in other organism | 0.00718 |
| GO:0044532 | modulation of apoptotic process in other organism | 0.00718 |
| GO:0052040 | modulation by symbiont of host programmed cell death | 0.00718 |
| GO:0052248 | modulation of programmed cell death in other organism involved in symbiotic interaction | 0.00718 |
| GO:0052433 | modulation by organism of apoptotic process in other organism involved in symbiotic interaction | 0.00718 |
| GO:0000117 | regulation of transcription involved in G2/M transition of mitotic cell cycle | 0.007233 |
| GO:0000738 | DNA catabolic process, exonucleolytic | 0.007283 |
| GO:0033692 | cellular polysaccharide biosynthetic process | 0.007979 |
| GO:0000245 | spliceosomal complex assembly | 0.008353 |
| GO:0035063 | nuclear speck organization | 0.008601 |
| GO:1904751 | positive regulation of protein localization to nucleolus | 0.009986 |
| GO:0044362 | negative regulation of molecular function in other organism | 0.010107 |
| GO:0052204 | negative regulation of molecular function in other organism involved in symbiotic interaction | 0.010107 |
| GO:1901216 | positive regulation of neuron death | 0.010112 |
| GO:0000271 | polysaccharide biosynthetic process | 0.010769 |
| GO:0016052 | carbohydrate catabolic process | 0.011035 |
| GO:1903525 | regulation of membrane tubulation | 0.011163 |
| GO:0034637 | cellular carbohydrate biosynthetic process | 0.011268 |
| GO:0051725 | protein de-ADP-ribosylation | 0.011689 |
| GO:2000035 | regulation of stem cell division | 0.011765 |

| | | |
|---|---|---|
| GO:0005977 | glycogen metabolic process | 0.012354 |
| GO:0006073 | cellular glucan metabolic process | 0.012354 |
| GO:0044042 | glucan metabolic process | 0.012354 |
| GO:0051168 | nuclear export | 0.012871 |
| GO:1904749 | regulation of protein localization to nucleolus | 0.012892 |
| GO:0001895 | retina homeostasis | 0.012915 |
| GO:0044359 | modulation of molecular function in other organism | 0.013661 |
| GO:0052205 | modulation of molecular function in other organism involved in symbiotic interaction | 0.013661 |
| GO:2001286 | regulation of caveolin-mediated endocytosis | 0.013777 |
| GO:0036462 | TRAIL-activated apoptotic signaling pathway | 0.013832 |
| GO:0008380 | RNA splicing | 0.014365 |
| GO:0005981 | regulation of glycogen catabolic process | 0.0144 |
| GO:0019318 | hexose metabolic process | 0.014588 |
| GO:0097202 | activation of cysteine-type endopeptidase activity | 0.01512 |
| GO:0043255 | regulation of carbohydrate biosynthetic process | 0.015464 |
| GO:0006285 | base-excision repair, AP site formation | 0.015553 |
| GO:0045008 | depyrimidination | 0.015553 |
| GO:0043696 | dedifferentiation | 0.015942 |
| GO:0043697 | cell dedifferentiation | 0.015942 |
| GO:1903729 | regulation of plasma membrane organization | 0.016129 |
| GO:2001237 | negative regulation of extrinsic apoptotic signaling pathway | 0.016191 |
| GO:0006112 | energy reserve metabolic process | 0.016881 |
| GO:0043470 | regulation of carbohydrate catabolic process | 0.017704 |
| GO:0071550 | death-inducing signaling complex assembly | 0.017706 |
| GO:0044264 | cellular polysaccharide metabolic process | 0.018014 |
| GO:0031125 | rRNA 3'-end processing | 0.018147 |
| GO:1902459 | positive regulation of stem cell population maintenance | 0.018827 |
| GO:1902570 | protein localization to nucleolus | 0.020274 |
| GO:0010906 | regulation of glucose metabolic process | 0.020515 |
| GO:0022613 | ribonucleoprotein complex biogenesis | 0.020953 |
| GO:0097296 | activation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway | 0.021103 |
| GO:2000650 | negative regulation of sodium ion transmembrane transporter activity | 0.021334 |
| GO:0006883 | cellular sodium ion homeostasis | 0.022645 |
| GO:0032025 | response to cobalt ion | 0.022887 |
| GO:0071803 | positive regulation of podosome assembly | 0.023113 |
| GO:0043471 | regulation of cellular carbohydrate catabolic process | 0.023152 |
| GO:0005976 | polysaccharide metabolic process | 0.023356 |

| | | |
|---|---|---|
| GO:0006913 | nucleocytoplasmic transport | 0.023804 |
| GO:0006304 | DNA modification | 0.024163 |
| GO:1902306 | negative regulation of sodium ion transmembrane transport | 0.024837 |
| GO:0051169 | nuclear transport | 0.024896 |
| GO:0005996 | monosaccharide metabolic process | 0.025243 |
| GO:0045651 | positive regulation of macrophage differentiation | 0.025616 |
| GO:2000009 | negative regulation of protein localization to cell surface | 0.026107 |
| GO:0019317 | fucose catabolic process | 0.027106 |
| GO:0042354 | L-fucose metabolic process | 0.027106 |
| GO:0042355 | L-fucose catabolic process | 0.027106 |
| GO:0051974 | negative regulation of telomerase activity | 0.028299 |
| GO:0009223 | pyrimidine deoxyribonucleotide catabolic process | 0.028442 |
| GO:0061000 | negative regulation of dendritic spine development | 0.028721 |
| GO:0072584 | caveolin-mediated endocytosis | 0.02935 |
| GO:0034643 | establishment of mitochondrion localization, microtubule-mediated | 0.030167 |
| GO:0047497 | mitochondrion transport along microtubule | 0.030167 |
| GO:0010766 | negative regulation of sodium ion transport | 0.030171 |
| GO:0006353, termination | DNA-templated transcription, termination | 0.030285 |
| GO:0006493 | protein O-linked glycosylation | 0.030557 |
| GO:0051654 | establishment of mitochondrion localization | 0.031091 |
| GO:0031123 | RNA 3'-end processing | 0.031568 |
| GO:2001269 | positive regulation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway | 0.032691 |
| GO:0043280 | positive regulation of cysteine-type endopeptidase activity involved in apoptotic process | 0.033074 |
| GO:0045649 | regulation of macrophage differentiation | 0.033708 |
| GO:0016071 | mRNA metabolic process | 0.034377 |
| GO:1901017 | negative regulation of potassium ion transmembrane transporter activity | 0.034678 |
| GO:0030575 | nuclear body organization | 0.034893 |
| GO:1901380 | negative regulation of potassium ion transmembrane transport | 0.035096 |
| GO:0006244 | pyrimidine nucleotide catabolic process | 0.0352 |
| GO:0010675 | regulation of cellular carbohydrate metabolic process | 0.035527 |
| GO:0071801 | regulation of podosome assembly | 0.035967 |
| GO:0006004 | fucose metabolic process | 0.036858 |
| GO:0051345 | positive regulation of hydrolase activity | 0.03714 |
| GO:0051302 | regulation of cell division | 0.037586 |
| GO:0045725 | positive regulation of glycogen biosynthetic process | 0.037904 |
| GO:0032211 | negative regulation of telomere maintenance via telomerase | 0.038012 |

| | | |
|---|---|---|
| GO:2001236 | regulation of extrinsic apoptotic signaling pathway | 0.038191 |
| GO:2001056 | positive regulation of cysteine-type endopeptidase activity | 0.038616 |
| GO:0090501 | RNA phosphodiester bond hydrolysis | 0.04007 |
| GO:0043467 | regulation of generation of precursor metabolites and energy | 0.040456 |
| GO:0060041 | retina development in camera-type eye | 0.040989 |
| GO:0009219 | pyrimidine deoxyribonucleotide metabolic process | 0.041415 |
| GO:0043628 | ncRNA 3'-end processing | 0.043752 |
| GO:0006486 | protein glycosylation | 0.043774 |
| GO:0043413 | macromolecule glycosylation | 0.043774 |
| GO:0016338 | calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules | 0.047051 |
| GO:0070875 | positive regulation of glycogen metabolic process | 0.047427 |
| GO:0034138 | toll-like receptor 3 signaling pathway | 0.047653 |
| GO:0070085 | glycosylation | 0.047803 |
| GO:0010977 | negative regulation of neuron projection development | 0.048276 |
| GO:0010950 | positive regulation of endopeptidase activity | 0.048306 |
| GO:1901099 | negative regulation of signal transduction in absence of ligand | 0.048537 |
| GO:2001240 | negative regulation of extrinsic apoptotic signaling pathway in absence of ligand | 0.048537 |
| GO:2000036 | regulation of stem cell population maintenance | 0.048947 |
| GO:0097320 | plasma membrane tubulation | 0.049208 |
| GO:0045821 | positive regulation of glycolytic process | 0.049741 |
| GO:0060544 | regulation of necroptotic process | 0.049919 |

**Table III-6: Summary of evidence for PGC-AUT GWAS results being meQTLs across 4 tissue types.** The number of ASD related (p-value <= 1E-5) SNPs and independent sites (defined via LD blocks) that fall under each category of cross-tissue meQTL. Results are shown for meQTLs associated at FDR = 1% threshold in blood, cord blood, and placenta datasets and threshold of 1E-8 in fetal brain. All SNPs in PGC, regardless of if they were tested in the respective meQTL studies, are included in this analysis. Independent sites were constructed by grouping SNPs into bins defined by recombination hot spot data from 1000 Genomes (see Materials and Methods). For example, scenario 1 lists that there are a total of 2173 SNPs that are meQTLs in blood, cord blood, and fetal brain, which fall into 16 loci (defined via LD blocks).

| Scenario | Blood | Cord Blood | Placenta | Fetal Brain | SNPs | % of Total SNPs | Independent Sites | % of Total Independent Sites |
|---|---|---|---|---|---|---|---|---|
| ***1*** | ✓ | ✓ | ✓ | ✓ | 2173 | 53.96% | 16 | 12.40% |
| ***2*** | ✓ | ✓ | ✓ | × | 90 | 2.23% | 10 | 7.75% |
| ***3*** | ✓ | × | ✓ | ✓ | 8 | 0.20% | 2 | 1.55% |
| ***4*** | ✓ | ✓ | × | ✓ | 1 | 0.02% | 0 | 0.00% |
| ***5*** | ✓ | ✓ | × | × | 0 | 0.00% | 0 | 0.00% |
| ***6*** | ✓ | × | ✓ | × | 69 | 1.71% | 9 | 6.98% |
| ***7*** | ✓ | × | × | ✓ | 29 | 0.72% | 0 | 0.00% |
| ***8*** | ✓ | × | × | × | 14 | 0.35% | 4 | 3.10% |
| ***9*** | × | ✓ | ✓ | ✓ | 684 | 16.99% | 2 | 1.55% |
| ***10*** | × | ✓ | ✓ | × | 88 | 2.19% | 4 | 3.10% |
| ***11*** | × | ✓ | × | ✓ | 3 | 0.07% | 0 | 0.00% |
| ***12*** | × | ✓ | × | × | 4 | 0.10% | 0 | 0.00% |
| ***13*** | × | × | ✓ | ✓ | 23 | 0.57% | 1 | 0.78% |
| ***14*** | × | × | ✓ | × | 593 | 14.73% | 35 | 27.13% |
| ***15*** | × | × | × | ✓ | 9 | 0.22% | 0 | 0.00% |
| ***16*** | × | × | × | × | 239 | 5.93% | 46 | 35.66% |
| | | | | **SUM** | 4027 | | 129 | |

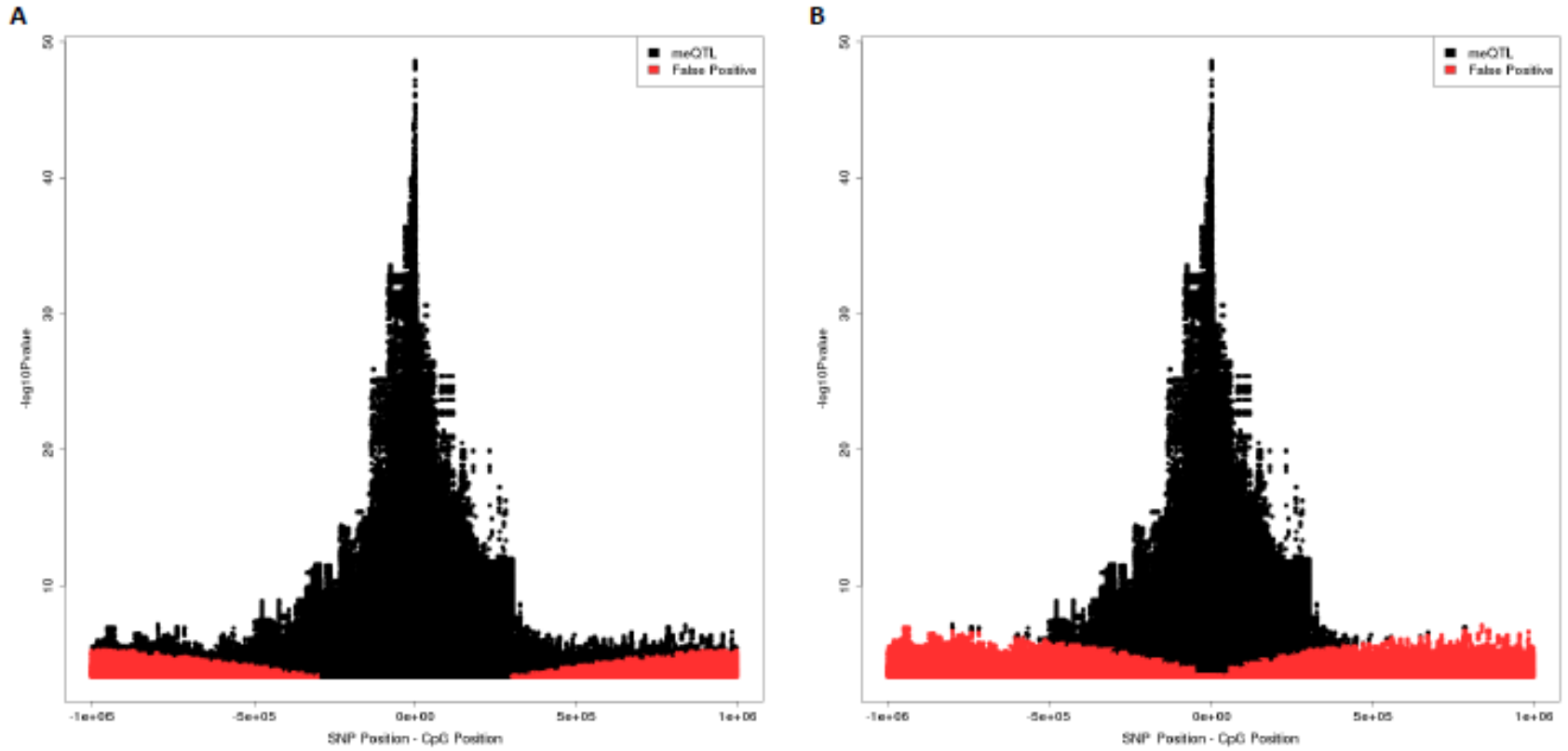



**Figure III-1: The relationship between significance level and SNP and CpG site distance on chromosome 21.** Degree of significance (y-axis) defined by -log10 p-value. The degree of significance decays with increasing distance. Points are colored by status as a meQTL or a false positive as determined by IHW (Ignatiadis, Klaus, Zaugg, & Huber, 2016). **Panel A**) FDR = 5%. **Panel B**) FDR = 1%.

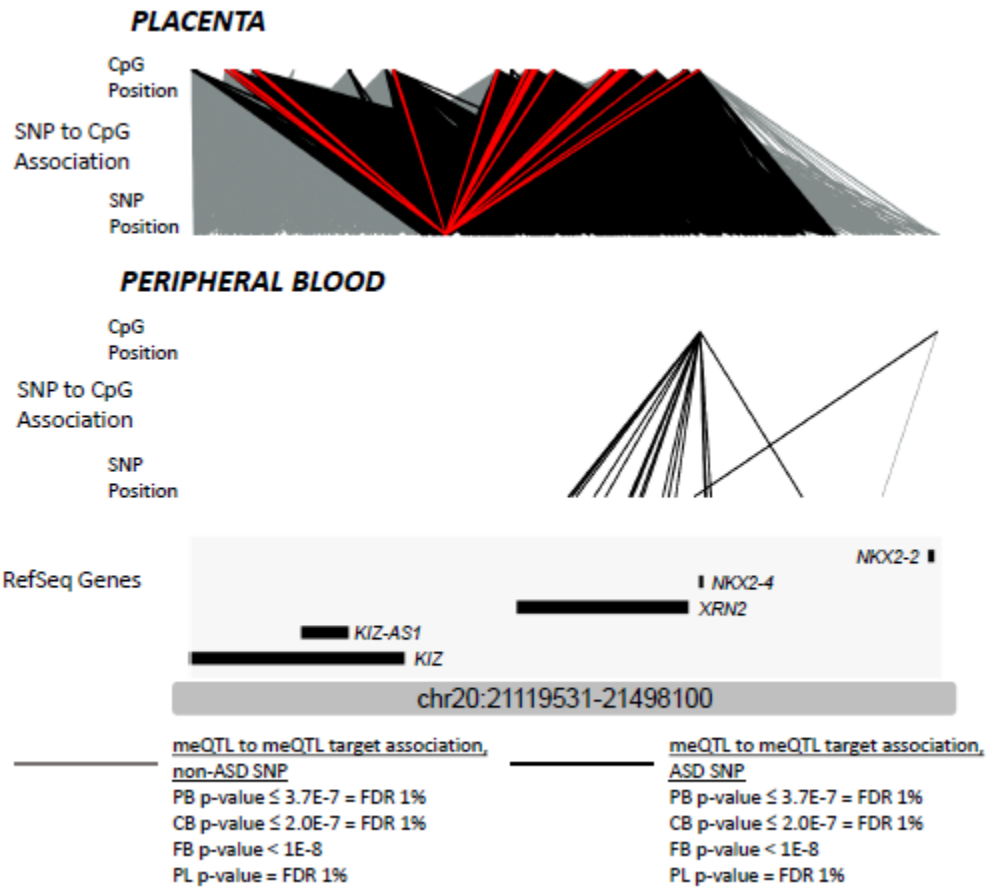



**Figure III-2: Identification of ASD candidate genes through meQTL mapping in peripheral blood and placenta for an ASD GWAS locus identified on chromosome 20.** Each tissue-specific panel presents, from bottom to top: genomic location, gene annotations, SNP locations, SNP-CpG associations, CpG locations. Light gray meQTL association lines denote all SNP to CpG associations in that tissue type. Dark meQTL association lines denote SNP-CpG associations for ASD-associated SNPs in PGC (*P*-value <= 1e-05). Red lines denote SNP-CpG associations for genome-wide significant SNP (rs910805) harbored in this locus. Data are presented for meQTL maps for (from top to bottom): placenta and peripheral blood. Please note locus coordinates differ from those in **Table 2.2** because in this context they encompass the locations of meQTL target CpG sites.

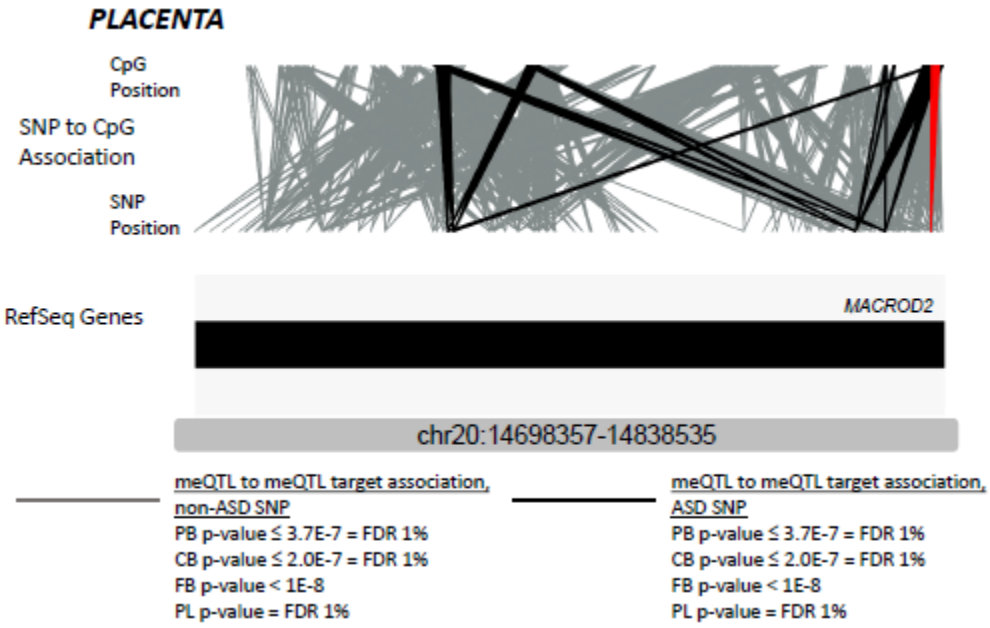



**Figure III-3: Mapping potential functional loci for follow up studies in placenta tissue for the PGC-AUT GWAS identified SNP at the *MACROD2* locus on chromosome 20.** The panels present, from bottom to top: genomic location, gene annotations, SNP locations, SNP-CpG associations, CpG locations. Light gray meQTL association lines denote all SNP to CpG associations in that tissue type. Dark meQTL association lines denote SNP-CpG associations for ASD-associated SNPs in PGC (*P*-value <= 1e-05). Red lines denote SNP-CpG associations for genome-wide significant SNP (rs71190156) harbored in this locus. Please note locus coordinates differ from those in **Table 2.2** because in this context they encompass the locations of meQTL target CpG sites.

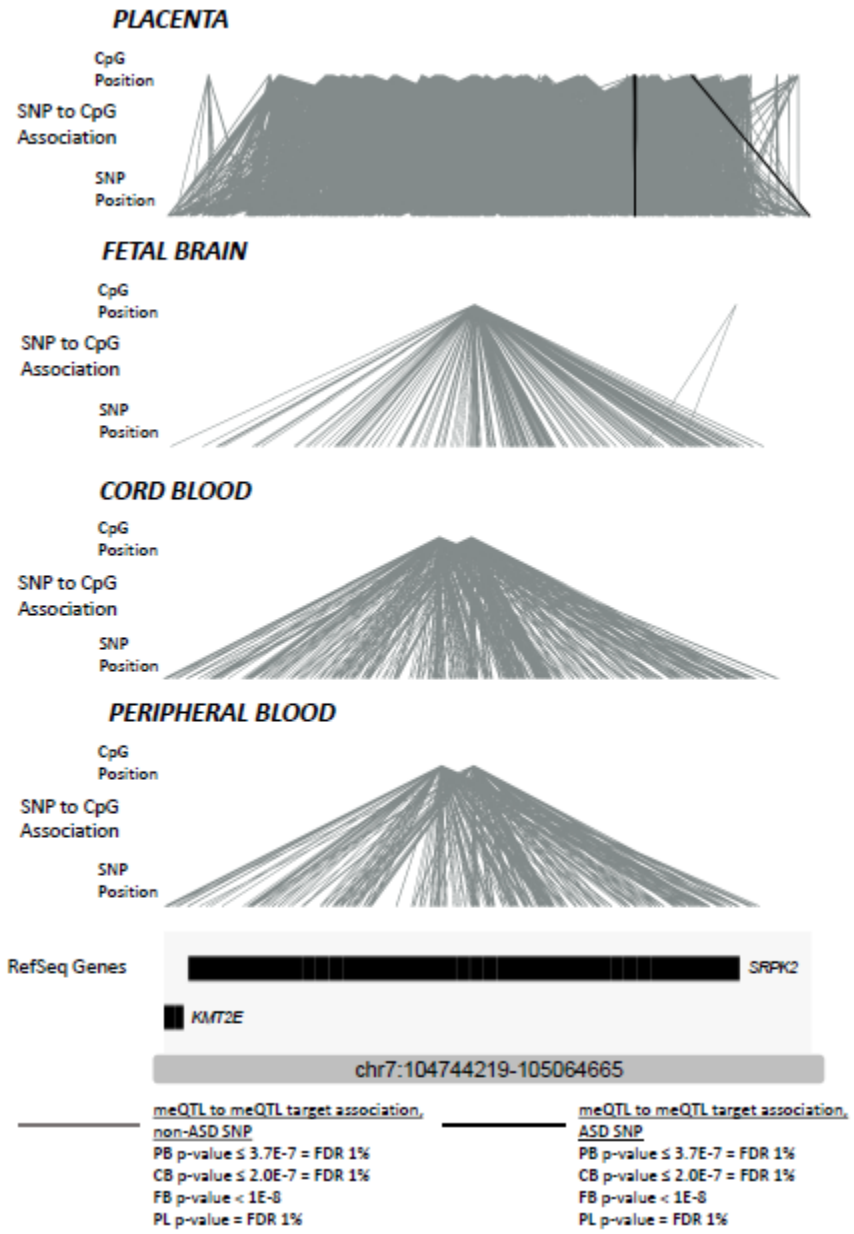



**Figure III-4: Mapping potential functional loci for follow up studies in placenta tissue for the PGC-AUT GWAS identified SNP at the *SRPK2* locus on chromosome 7.** The panels show, from bottom to top: genomic location, gene annotations, SNP locations, SNP-CpG associations, CpG locations. Light gray meQTL association lines denote all SNP to CpG associations in that tissue type. Dark meQTL association lines denote SNP-CpG associations for ASD-associated SNPs in PGC (*P*-value <= 1e-05). Please note locus coordinates differ from those in **Table 2.2** because in this context they encompass the locations of meQTL target CpG sites.

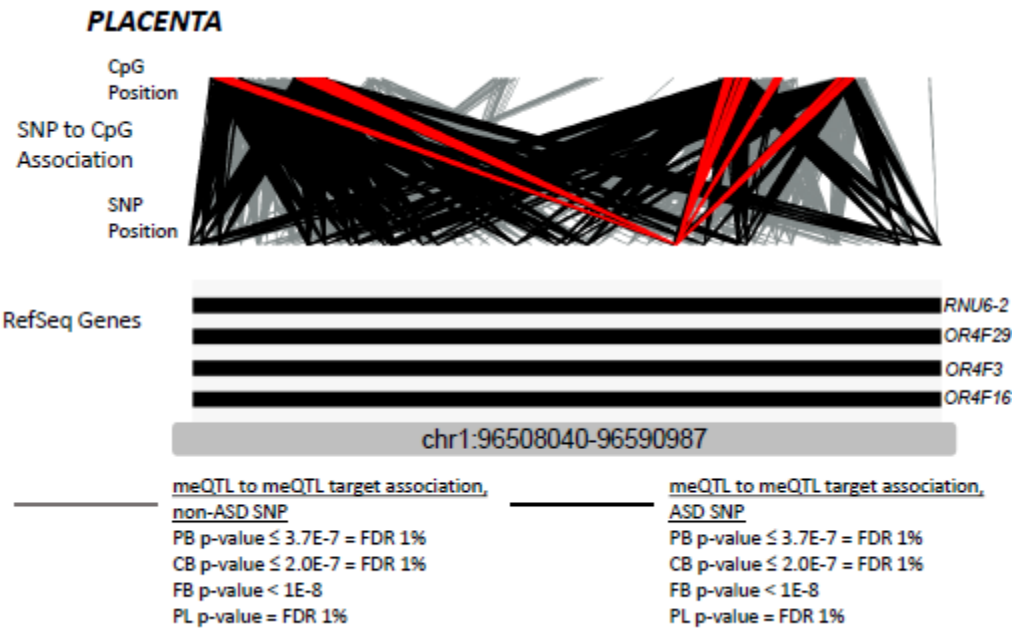



**Figure III-5: Mapping potential functional loci for follow up studies in placenta tissue for the PGC-AUT GWAS identified SNP at a genomic region on chromosome 1.** The panels present, from bottom to top: genomic location, gene annotations, SNP locations, SNP-CpG associations, CpG locations. Light gray meQTL association lines denote all SNP to CpG associations in that tissue type. Dark meQTL association lines denote SNP-CpG associations for ASD-associated SNPs in PGC (*P*-value <= 1e-05). Red lines denote SNP-CpG associations for genome-wide significant SNP (rs201910565) harbored in this locus. Please note locus coordinates differ from those in **Table 2.2** because they encompass the locations of meQTL target CpG sites.

# APPENDIX IV

**Table IV-1: Demographics table for WGBS samples included in ASD DMR and global methylation analyses.**

| | Males (n = 61) | | | p-value | Females (n = 45) | | | p-value |
|---|---|---|---|---|---|---|---|---|
| | **ASD (n = 18)** | **Non-TD (n = 26)** | **TD (n = 17)** | | **ASD (n = 1)** | **Non-TD (n = 24)** | **TD (n = 20)** | |
| ***Gestational Age*** | **Mean (range)** | | | 0.422[a,1]; 0.879[a,2] | **Mean (range)** | | | 0.821[a,2] |
| Weeks | 39.05 (35.99 - 41.00) | 39.36 (36.71 - 42.43) | 39.43 (36.29 - 42.43) | | 40.01 | 39.22 (34.28 - 41.29) | 39.12 (33.86 - 41.28) | |
| ***Race*** | **N (%)** | | | 0.275[b,1]; 0.126[b,2] | **N (%)** | | | 0.00659[b,2] |
| White | 11 (0.61) | 12 (0.46) | 12 (0.71) | | 0 | 6 (0.25) | 13 (0.65) | |
| Black | 5 (0.28) | 6 (0.23) | 1 (0.06) | | 0 | 5 (0.21) | 0 | |
| Asian | 0 | 4 (0.15) | 0 | | 0 | 1 (0.04) | 2 (0.1) | |
| Other/Missing | 2 (0.11) | 4 (0.15) | 4 (0.24) | | 1 (1.00) | 12 (0.5) | 5 (0.25) | |
| ***Mode of delivery*** | **N (%)** | | | 0.638[b,1]; 0.837[b,2] | **N (%)** | | | 0.03[b,2] |
| Vaginal | 8 (0.44) | 15 (0.58) | 9 (0.53) | | 0 | 12 (0.5) | 14 (0.70) | |
| C-section | 7 (0.39) | 10 (0.38) | 4 (0.24) | | 0 | 10 (0.42) | 1 (0.05) | |
| Missing | 3 (0.17) | 1 (0.04) | 4 (0.24) | | 1 (1.00) | 2 (0.08) | 5 (0.25) | |

a: t-test, b: chi-square test, 1: ASD to TD comparison, 2: NTD to TD comparison

**Table IV-2: Summary statistics for sex-stratified diagnostic groups for region-specific global methylation values**

| Islands | | | | | | |
|---|---|---|---|---|---|---|
| | ***Minimum*** | ***1st Quartile*** | ***Median*** | ***Mean*** | ***3rd Quartile*** | ***Maximum*** |
| ***TD-Males*** | 0.211 | 0.221 | 0.223 | 0.224 | 0.225 | 0.255 |
| ***NTD-Males*** | 0.199 | 0.214 | 0.221 | 0.221 | 0.227 | 0.241 |
| ***ASD-Males*** | 0.211 | 0.222 | 0.226 | 0.226 | 0.232 | 0.235 |
| ***TD-Females*** | 0.210 | 0.219 | 0.222 | 0.223 | 0.225 | 0.238 |
| ***NTD-Females*** | 0.210 | 0.217 | 0.222 | 0.222 | 0.225 | 0.238 |
| **Shores[a]** | | | | | | |
| | ***Minimum*** | ***1st Quartile*** | ***Median*** | ***Mean*** | ***3rd Quartile*** | ***Maximum*** |
| ***TD-Males*** | 0.583 | 0.590 | 0.595 | 0.596 | 0.600 | 0.619 |
| ***NTD-Males[1]*** | 0.564 | 0.585 | 0.592 | 0.590 | 0.598 | 0.604 |
| ***ASD-Males*** | 0.583 | 0.592 | 0.597 | 0.596 | 0.599 | 0.609 |
| ***TD-Females*** | 0.565 | 0.583 | 0.591 | 0.590 | 0.598 | 0.606 |
| ***NTD-Females*** | 0.573 | 0.584 | 0.590 | 0.590 | 0.595 | 0.602 |
| **Shelves[b]** | | | | | | |
| | ***Minimum*** | ***1st Quartile*** | ***Median*** | ***Mean*** | ***3rd Quartile*** | ***Maximum*** |
| ***TD-Males*** | 0.707 | 0.712 | 0.717 | 0.718 | 0.724 | 0.735 |
| ***NTD-Males[2]*** | 0.684 | 0.707 | 0.715 | 0.712 | 0.719 | 0.731 |
| ***ASD-Males*** | 0.707 | 0.711 | 0.714 | 0.717 | 0.724 | 0.728 |
| ***TD-Females*** | 0.682 | 0.704 | 0.714 | 0.713 | 0.724 | 0.732 |
| ***NTD-Females*** | 0.689 | 0.706 | 0.713 | 0.712 | 0.720 | 0.728 |
| **Open Sea[c]** | | | | | | |
| | ***Minimum*** | ***1st Quartile*** | ***Median*** | ***Mean*** | ***3rd Quartile*** | ***Maximum*** |
| ***TD-Males*** | 0.602 | 0.612 | 0.617 | 0.618 | 0.623 | 0.640 |
| ***NTD-Males[3]*** | 0.555 | 0.602 | 0.611 | 0.608 | 0.617 | 0.632 |
| ***ASD-Males*** | 0.597 | 0.609 | 0.613 | 0.615 | 0.620 | 0.632 |
| ***TD-Females*** | 0.573 | 0.600 | 0.619 | 0.613 | 0.626 | 0.642 |
| ***NTD-Females*** | 0.584 | 0.602 | 0.612 | 0.612 | 0.622 | 0.637 |

a: 2000 bp downstream and upstream of islands; b: 2000 bp downstream and upstream of shores; c: measured CpG sites not in islands, shores, or shelves 1: nominally different from TD-males (t-test p-value = 0.043); 2: nominally different from TD-males (t-test p-value = 0.037); 3: nominally different from TD-males (t-test p-value = 0.014)

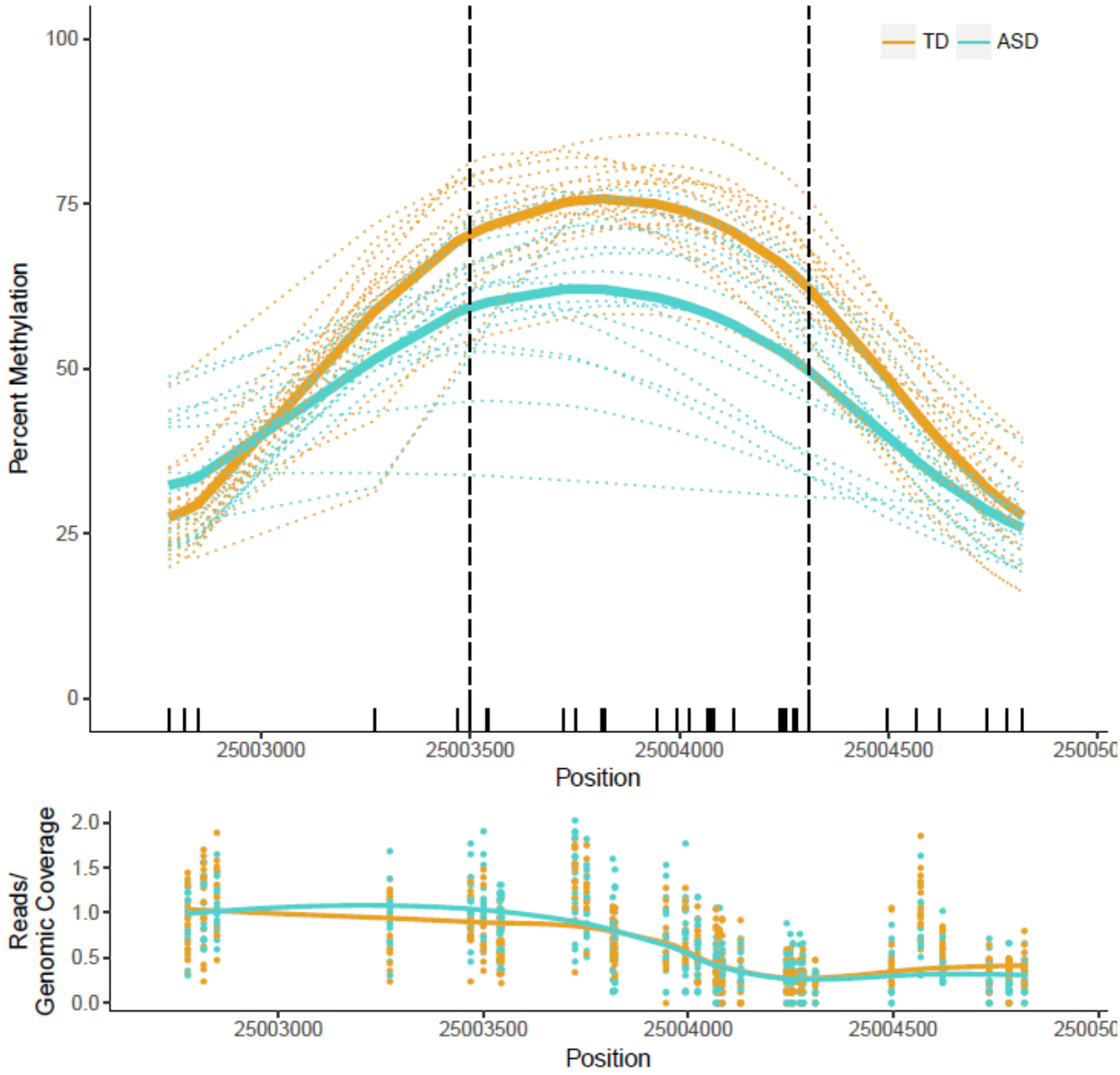



**Figure IV-1: Top-ranked DMR for comparison of ASD (n = 18) to TD (n = 17) in males**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Bottom panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis.

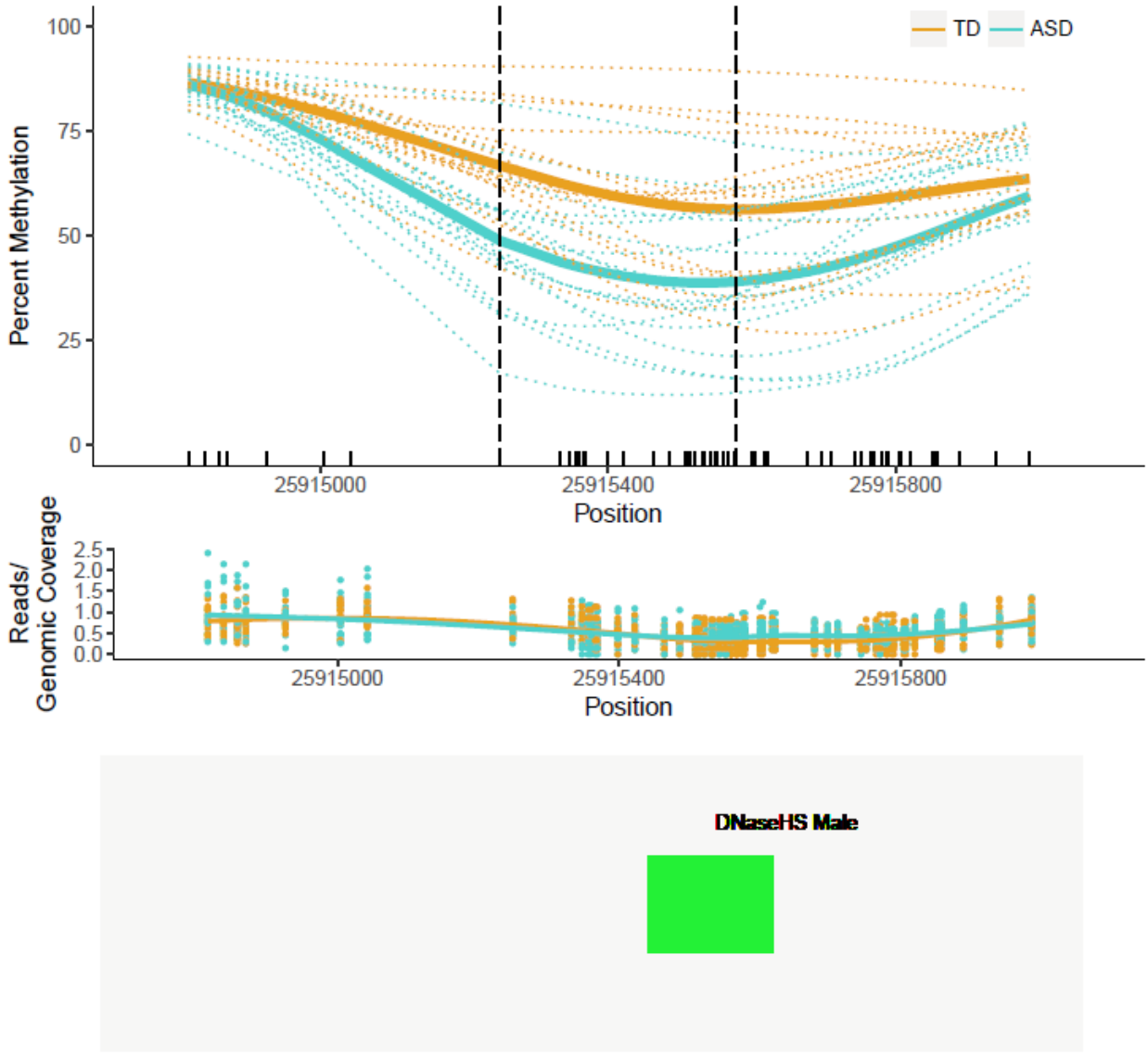



**Figure IV-2: Second-ranked DMR for comparison of ASD (n = 18) to TD (n = 17) in males**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

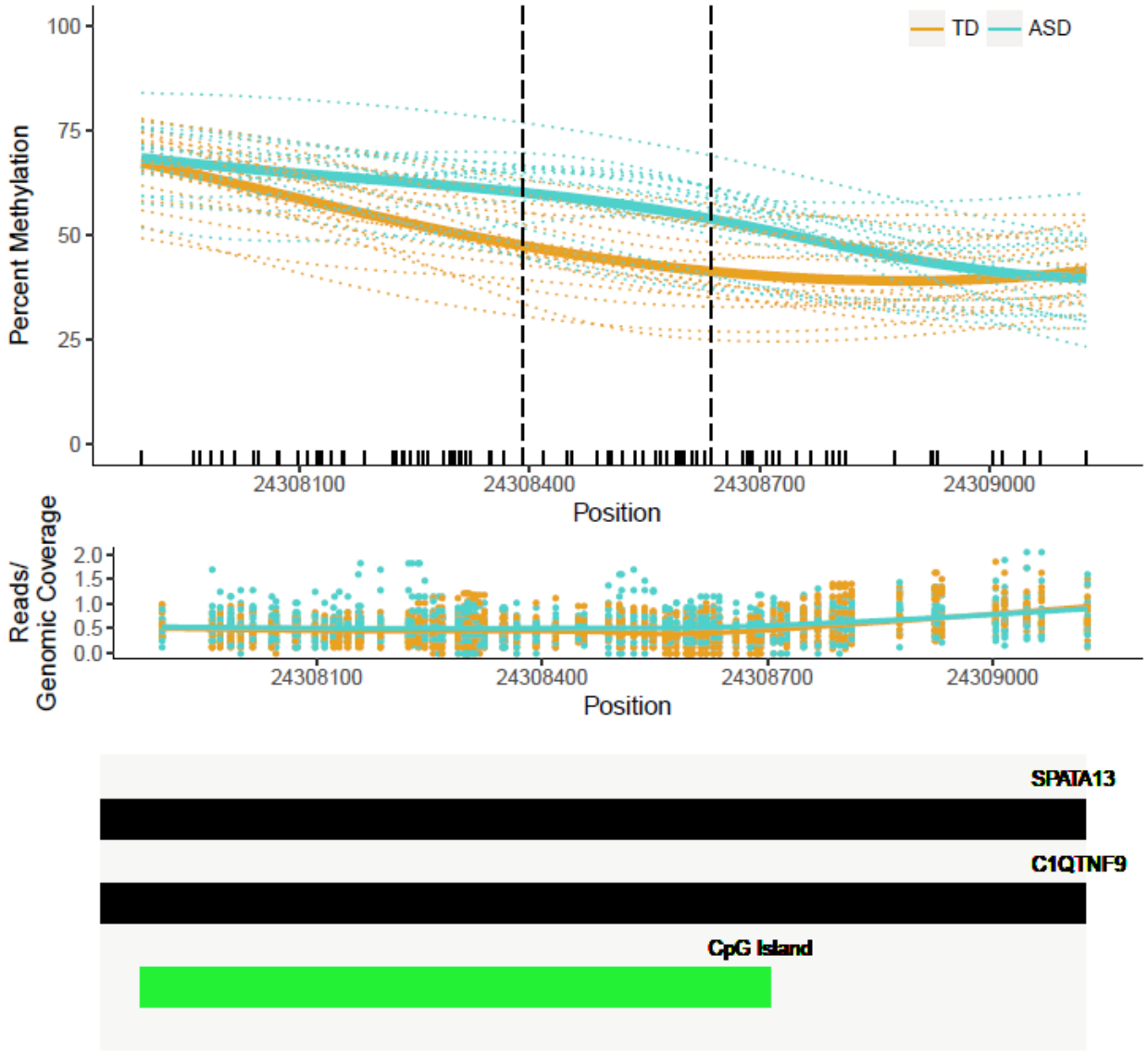



**Figure IV-3: Third-ranked DMR for comparison of ASD (n = 18) to TD (n = 17) in males**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

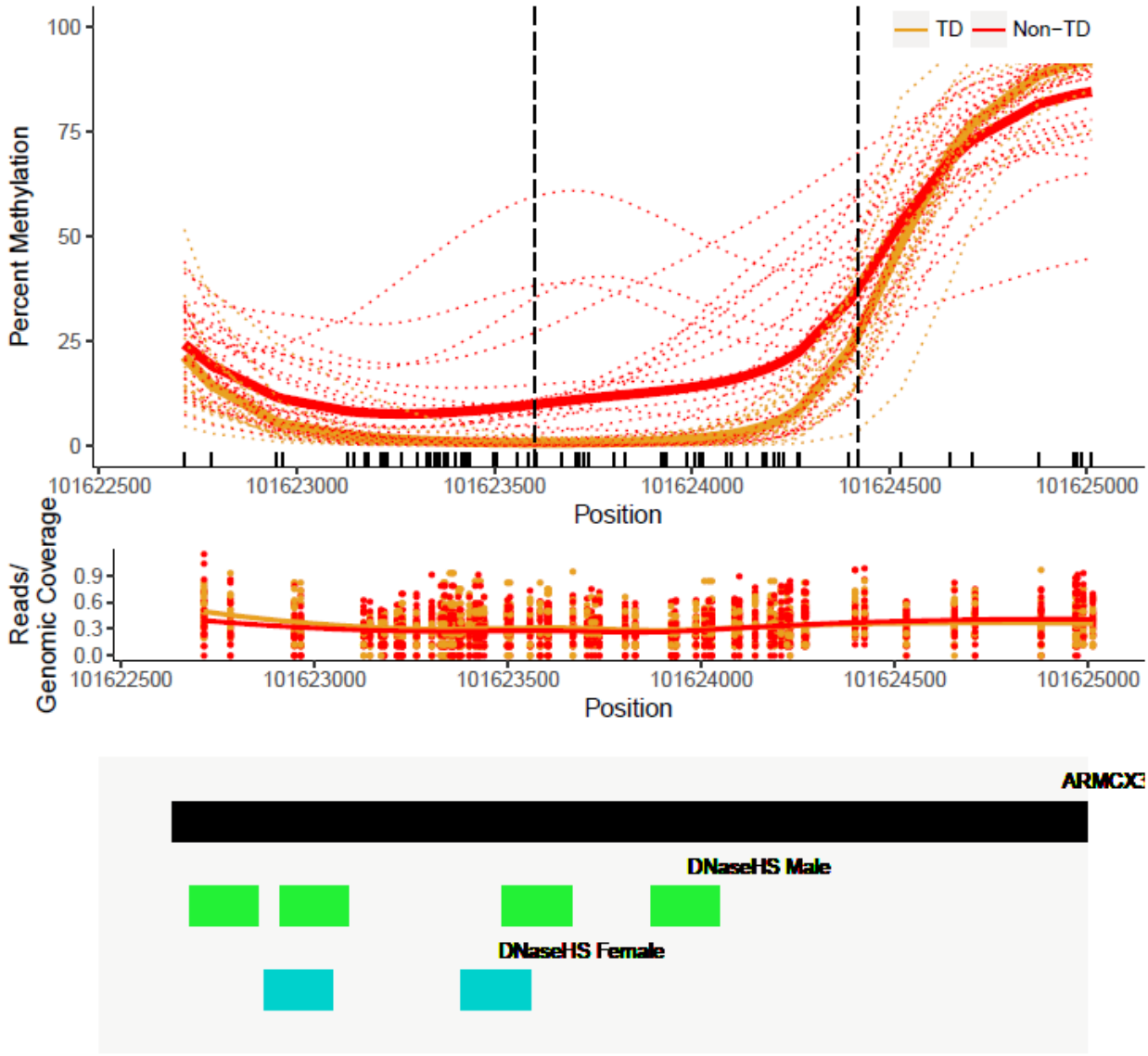



**Figure IV-4: Top-ranked DMR for comparison of NTD (n = 26) to TD (n = 17) in males**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

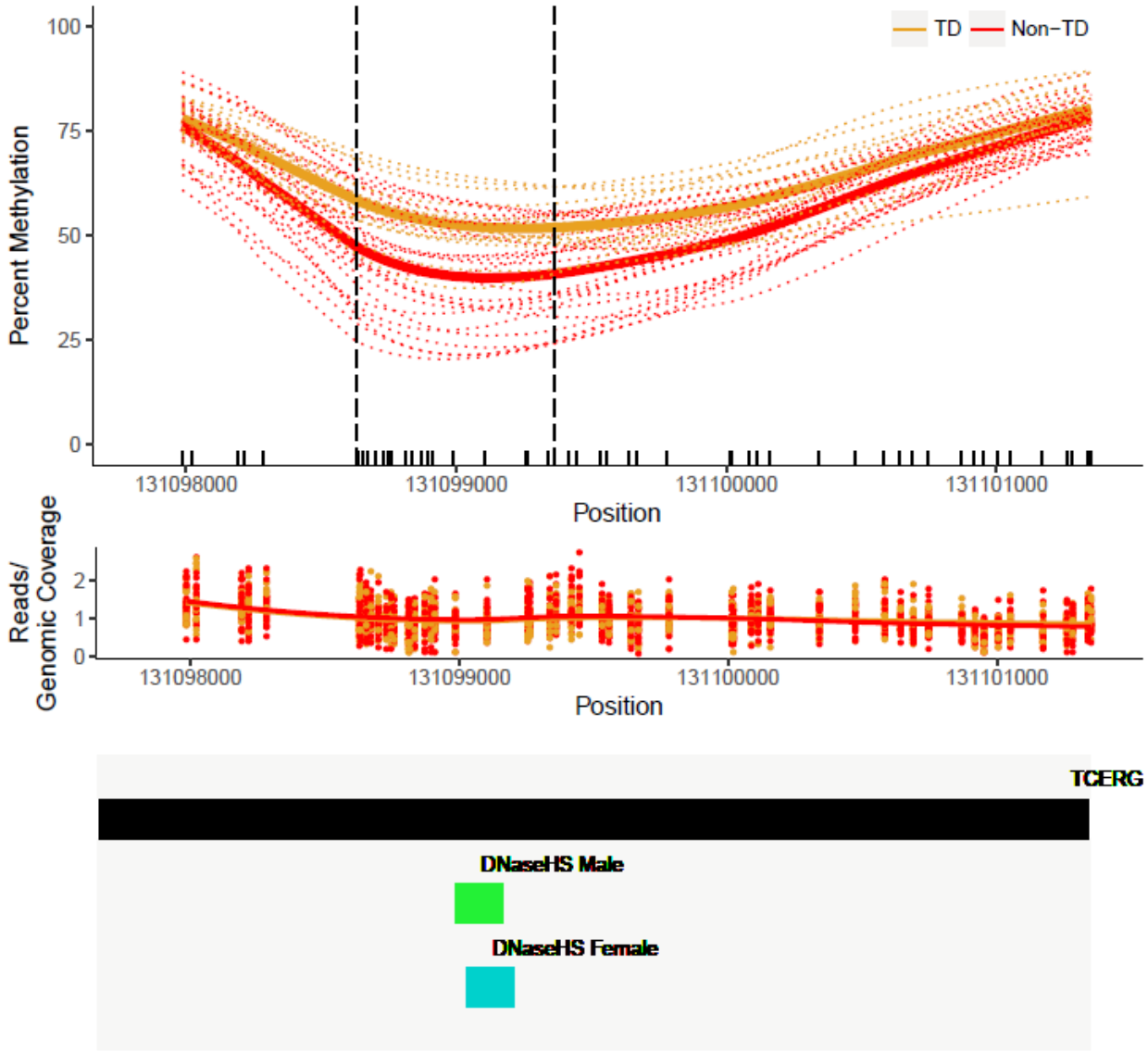



**Figure IV-5: Second-ranked DMR for comparison of NTD (n = 26) to TD (n = 17) in males**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

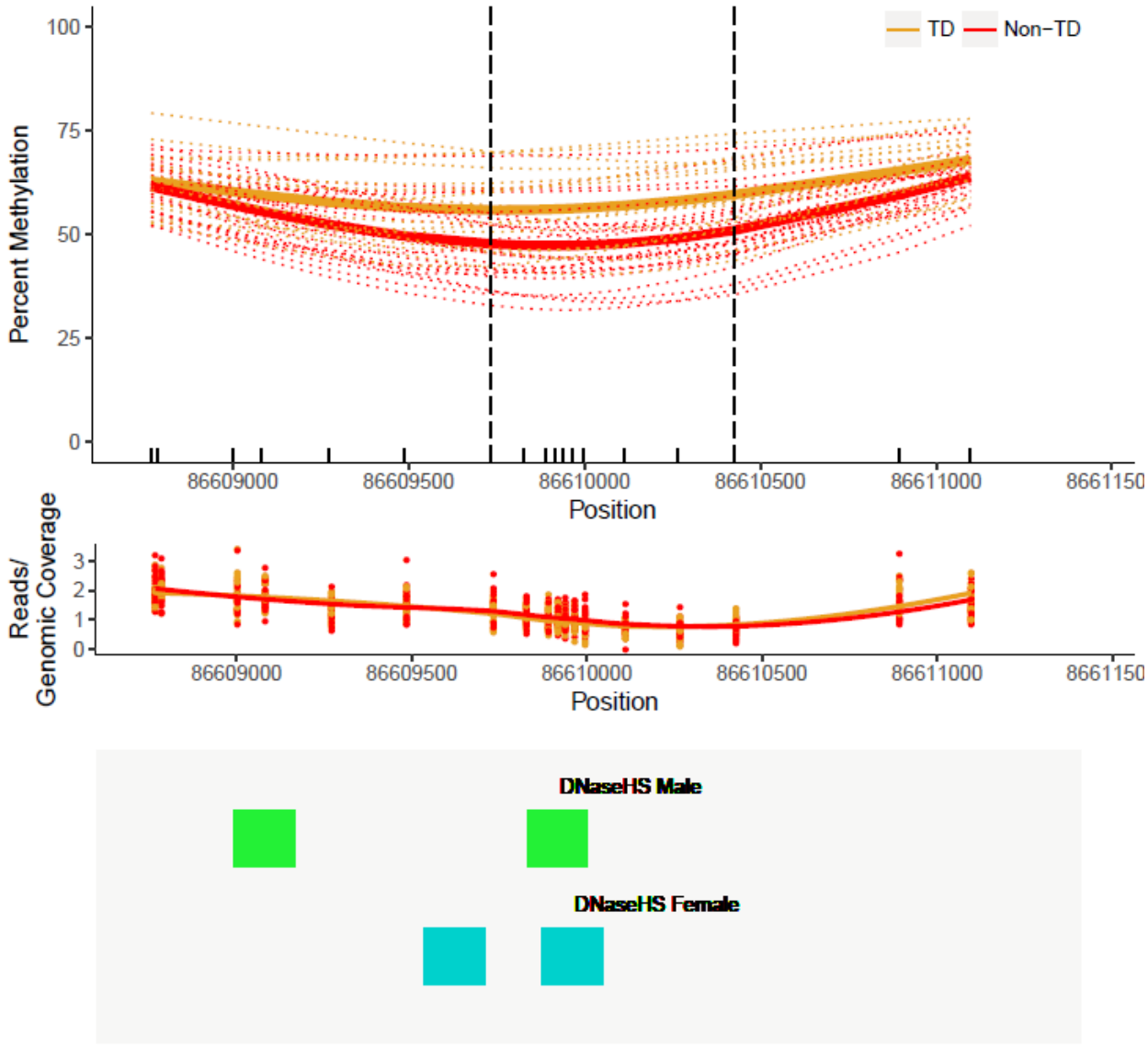



**Figure IV-6: Third-ranked DMR for comparison of NTD (n = 26) to TD (n = 17) in males**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

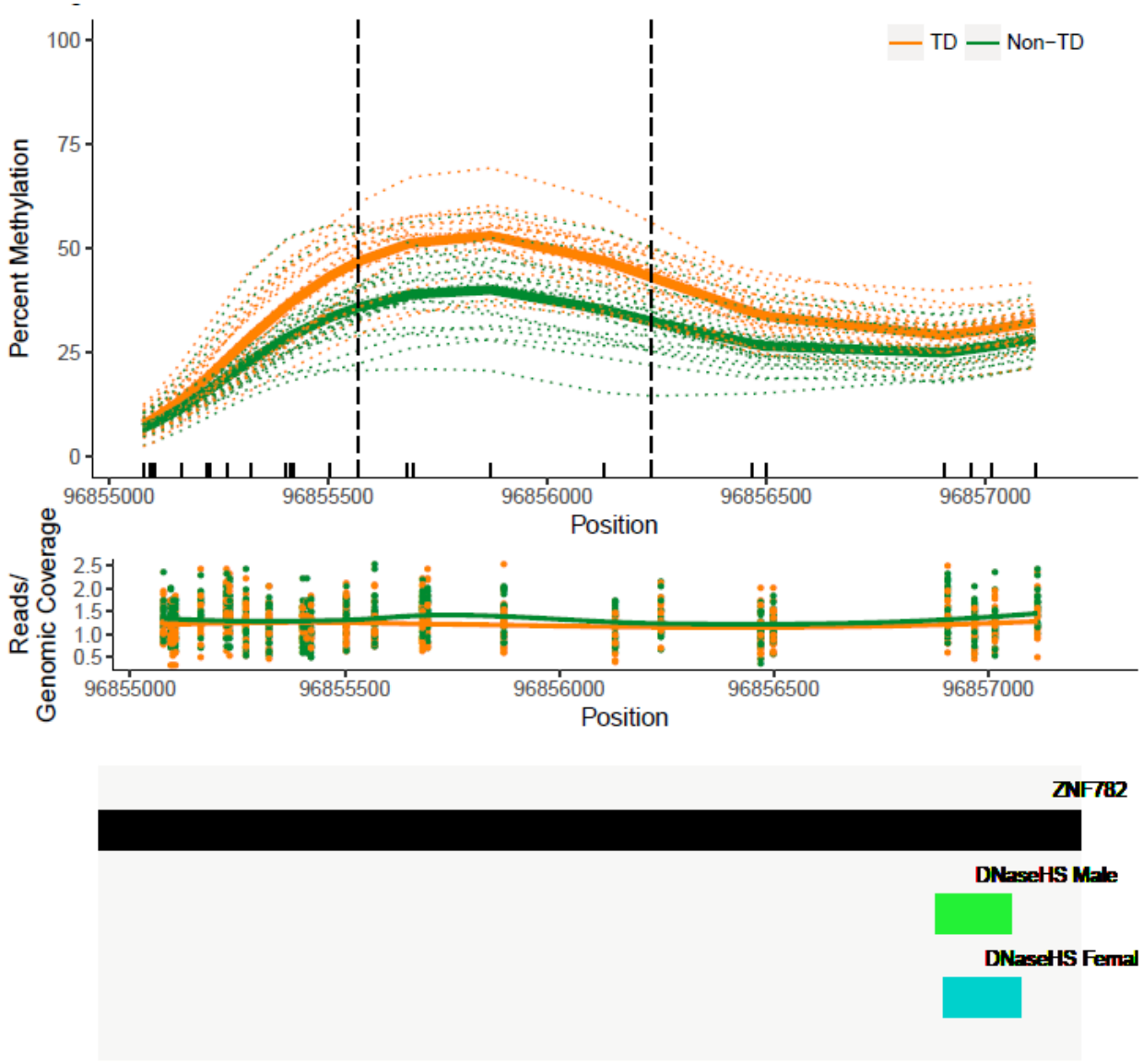



**Figure IV-7: Top-ranked DMR for comparison of NTD (n = 24) to TD (n = 20) in females**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

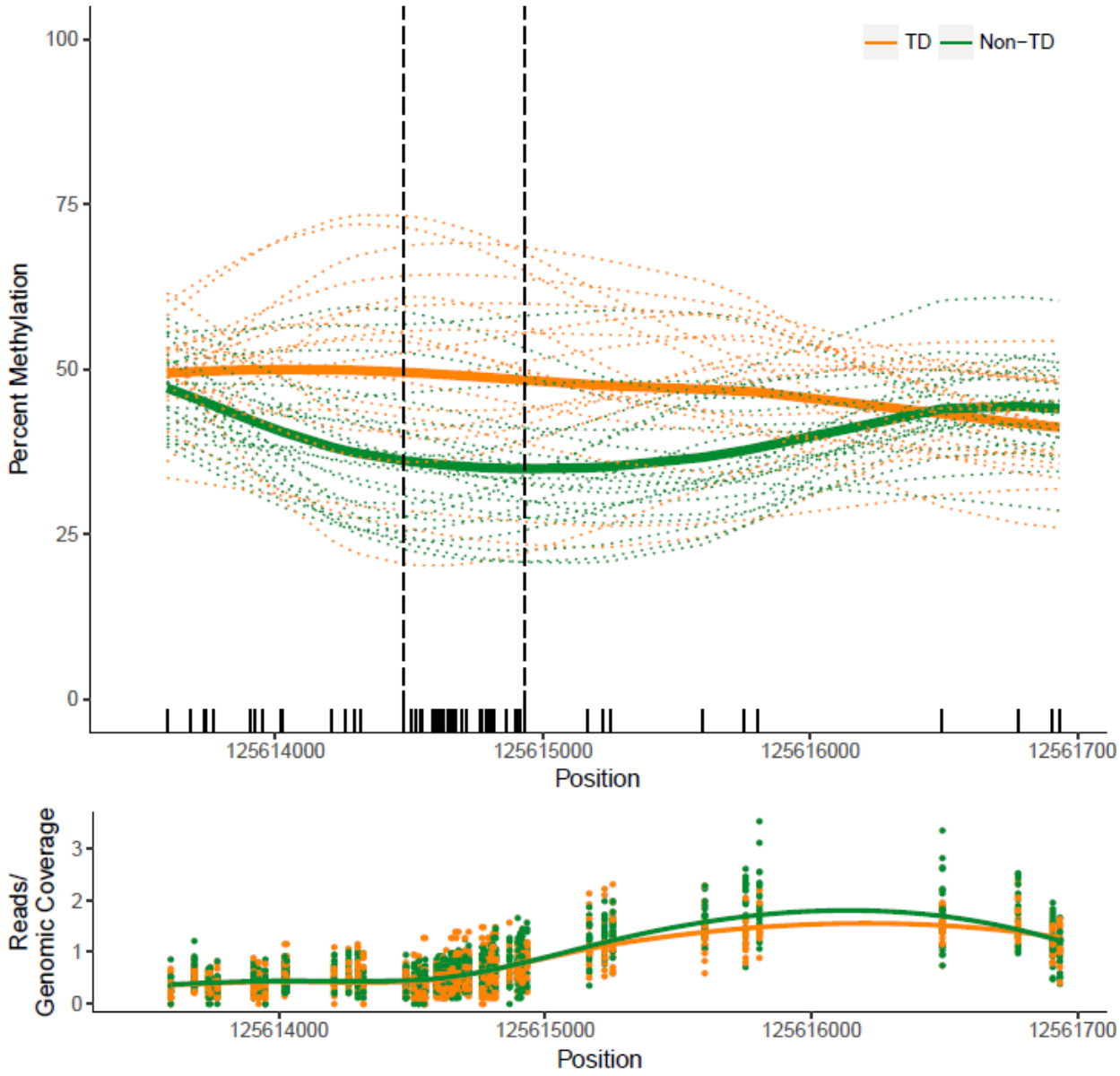



**Figure IV-8: Second-ranked DMR for comparison of NTD (n = 24) to TD (n = 20) in females**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Bottom panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis.

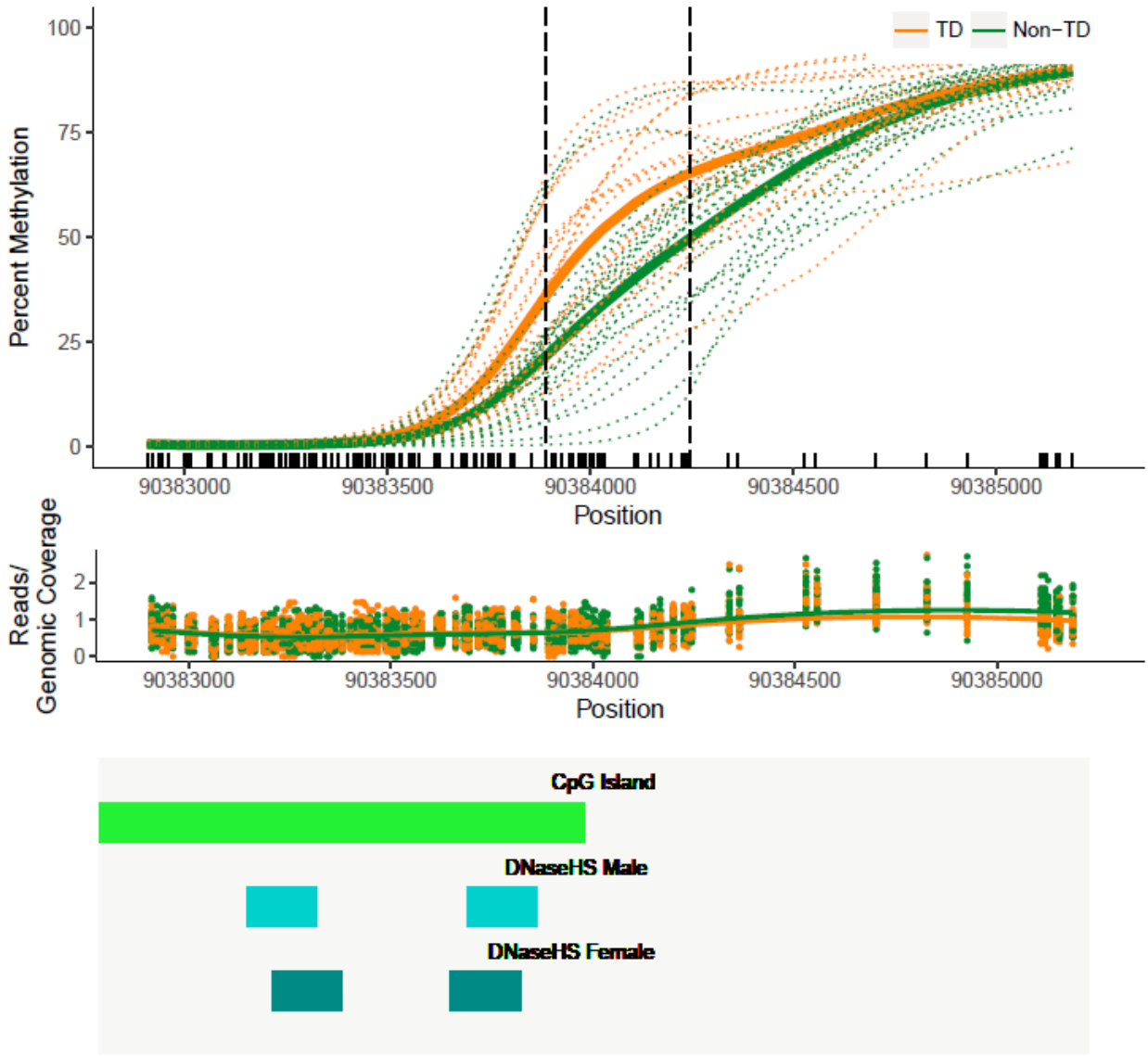



**Figure IV-9: Third-ranked DMR for comparison of NTD (n = 24) to TD (n = 20) in females**. Top panel: Percent methylation vs genomic position. Dashed black vertical lines represent actual DMR coordinates. Dashed lines represent individual samples and solid lines represent the average of samples of each respective diagnosis. Middle panel: Y-axis is ratio of reads at measured CpG sites divided by average coverage across the whole genome, x-axis is genomic position. This plot explores possibility of copy number differences by diagnosis. Bottom panel: Overlap of DMR with Gene track, CpG islands (UCSC Table Brower) and DNaseHS sites. Latter regions were downloaded from ENCODE for a male (GSM2400398) and female (GSM2400399) placenta sample separately.

# APPENDIX V

**Supplementary Table V-1: Sample characteristics for all 133 EARLI WGBS samples**

| | **Males (n=76)** | **Females (n=57)** | p-value |
|---|---|---|---|
| ***Diagnosis*** | ***N (%)*** | ***N (%)*** | 0.004[1] |
| Typically Developing | 17 (0.22) | 20 (0.35) | |
| Non-Typically Developing | 26 (0.34) | 24 (0.42) | |
| Autism Spectrum Disorder | 18 (0.24) | 1 (0.02) | |
| Lost to follow up | 15 (0.20) | 12 (0.21) | |
| ***Gestational Age*** | ***Mean (range)*** | ***Mean (range)*** | 0.604[2] |
| Weeks | 39.28 (35.99 - 42.43) | 39.41 (33.86 - 41.57) | |
| ***Race*** | ***N (%)*** | ***N (%)*** | 0.467[1] |
| White | 43 (0.57) | 26 (0.46) | |
| Black | 6 (0.08) | 4 (0.07) | |
| Asian | 12 (0.16) | 8 (0.14) | |
| Other/Missing | 15 (0.20) | 19 (0.33) | |
| ***Mode of delivery*** | ***N (%)*** | ***N (%)*** | 0.290[1] |
| Vaginal | 38 (0.50) | 30 (0.53) | |
| C-section | 26 (0.34) | 12 (0.21) | |
| Missing | 12 (0.16) | 15 (0.26) | |

[1]: chi-square test p-value

[2]: t-test p-value

**Supplementary Table V-2: BLAT results for *ZNF300* DMR region.** Results include BLAT results for reference sequence in DMR region and for source sequence of 9 450k probes that overlap this region.

450k Probes Overlapping *ZNF300 DMR*

| ACTIONS | QUERY | SCORE | START | END | QSIZE | IDENTITY | CHRO | STRAND | START | END | SPAN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| browser details | cg19486756 | 50 | 1 | 50 | 50 | 100.00% | 5 | - | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg12346881 | 50 | 1 | 50 | 50 | 100.00% | 5 | - | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg04675542 | 50 | 1 | 50 | 50 | 100.00% | 5 | + | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg02343823 | 50 | 1 | 50 | 50 | 100.00% | 5 | - | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg08580836 | 50 | 1 | 50 | 50 | 100.00% | 5 | - | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg08580836 | 32 | 12 | 46 | 50 | 97.10% | 6 | + | 1086171 | 1200162 | 113992 |
| browser details | cg19014419 | 50 | 1 | 50 | 50 | 100.00% | 5 | + | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg11291313 | 50 | 1 | 50 | 50 | 100.00% | 5 | - | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg18237551 | 50 | 1 | 50 | 50 | 100.00% | 5 | - | 1.5E+08 | 1.5E+08 | 50 |
| browser details | cg21228005 | 50 | 1 | 50 | 50 | 100.00% | 5 | + | 1.5E+08 | 1.5E+08 | 50 |

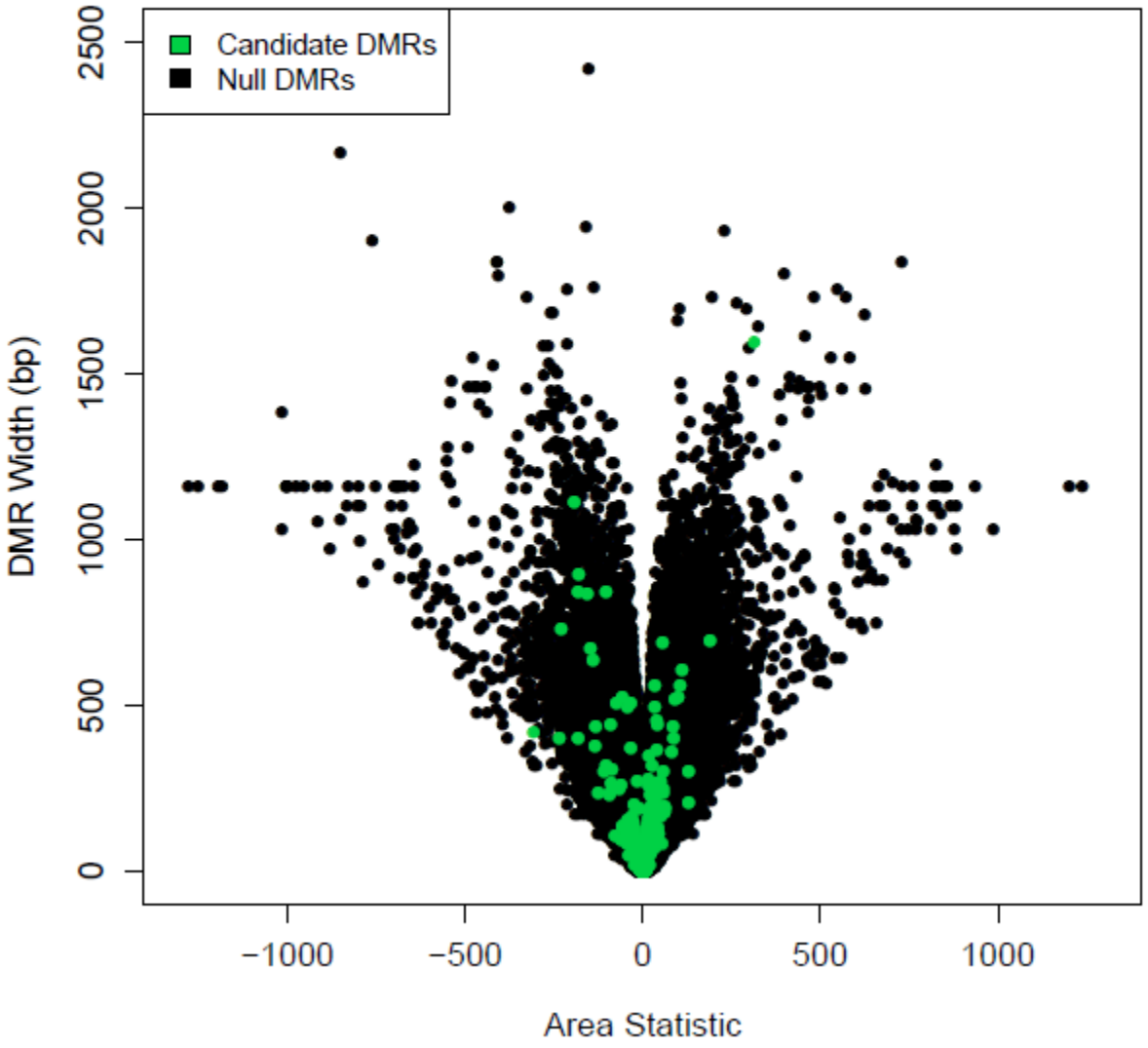
chr5:150904274-150905870 under hg38

| ACTIONS | QUERY | SCORE | START | END | QSIZE | IDENTITY | CHRO | STRAND | START | END | SPAN |
|---|---|---|---|---|---|---|---|---|---|---|---|
| browser details | hg38_dna | 1597 | 1 | 1597 | 1597 | 100.00% | 5 | + | 1.51E+08 | 1.51E+08 | 1597 |
| browser details | hg38_dna | 269 | 463 | 869 | 1597 | 83.50% | 5 | + | 1.51E+08 | 1.51E+08 | 403 |
| browser details | hg38_dna | 30 | 1170 | 1254 | 1597 | 96.90% | 1 | + | 1.54E+08 | 1.54E+08 | 233 |
| browser details | hg38_dna | 28 | 1199 | 1244 | 1597 | 66.70% | 1 | + | 50659355 | 50659384 | 30 |
| browser details | hg38_dna | 26 | 1352 | 1381 | 1597 | 96.50% | 10 | + | 461616 | 461660 | 45 |
| browser details | hg38_dna | 26 | 1352 | 1381 | 1597 | 96.50% | 10 | + | 746531 | 746575 | 45 |
| browser details | hg38_dna | 26 | 810 | 836 | 1597 | 100.00% | 1 | + | 1.94E+08 | 1.94E+08 | 483 |
| browser details | hg38_dna | 23 | 1182 | 1204 | 1597 | 100.00% | 11 | - | 12070379 | 12070401 | 23 |
| browser details | hg38_dna | 23 | 1510 | 1533 | 1597 | 100.00% | 1 | + | 1.76E+08 | 1.76E+08 | 25 |

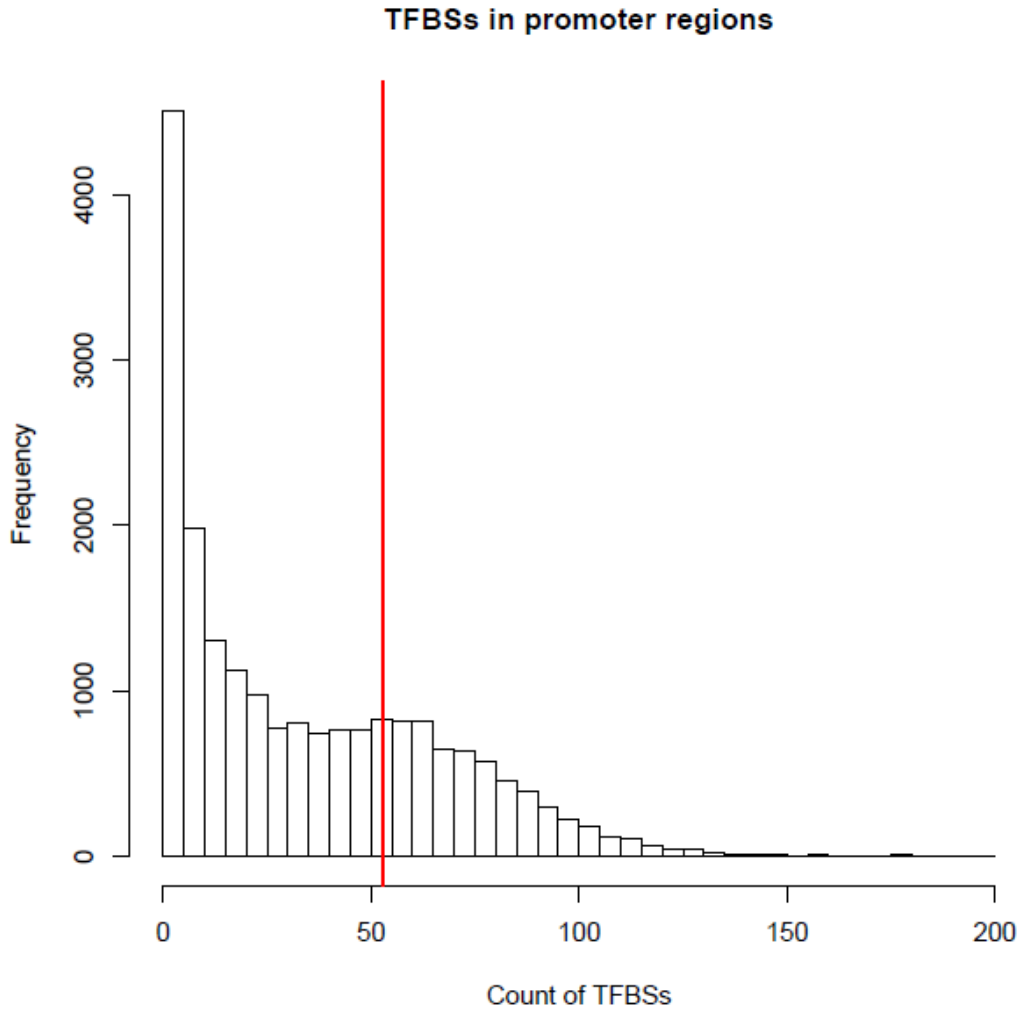
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| browser details | hg38_dna | 22 | 405 | 426 | 1597 | 100.00% | 7 | - | 37414236 | 37414257 | 22 |
| browser details | hg38_dna | 22 | 1182 | 1203 | 1597 | 100.00% | 12 | - | 1E+08 | 1E+08 | 22 |
| browser details | hg38_dna | 22 | 244 | 265 | 1597 | 100.00% | 1 | - | 22907061 | 22907082 | 22 |
| browser details | hg38_dna | 21 | 182 | 202 | 1597 | 100.00% | 12 | - | 70460496 | 70460516 | 21 |
| browser details | hg38_dna | 20 | 1182 | 1203 | 1597 | 95.50% | 1 | - | 29551036 | 29551057 | 22 |

**Supplementary Table V-3: Association metrics for mean *ZNF300* methylation levels and potential key confounders.** P-values shown are for linear regression of mean across 45 CpG sites in *ZNF300* DMR vs gestational age, the first 5 principal components from genotype data, and mode of delivery (vaginal vs C-section).

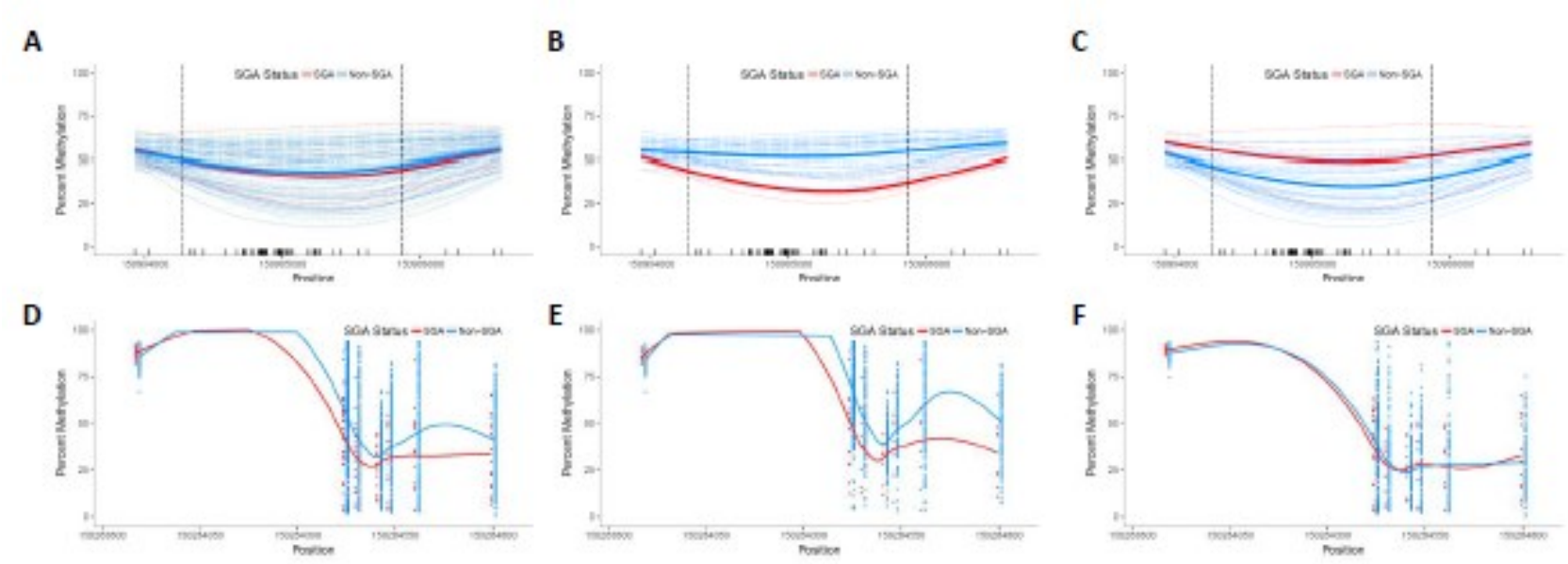
| $E(M_{DMRmean})= \alpha+ \beta_1 GestationalAge+\varepsilon$ | | |
|---|---|---|
| | *Coefficient (95% Confidence Interval)* | *p-value* |
| Gestational Age | 0.0025 (-0.0260, 0.0309) | 0.861 |
| $E(M_{DMRmean})= \alpha+ \beta_2 PC1+ \beta_2 PC2+ \beta_3 PC3+ \beta_4 PC4+ \beta_5 PC5+\varepsilon$ | | |
| | *Coefficient (95% Confidence Interval)* | *p-value* |
| PC1 | -1.52 (-7.95, 4.91) | 0.631 |
| PC2 | -2.09 (-9.05, 4.87) | 0.541 |
| PC3 | 1.29 (-12.34, 14.93) | 0.846 |
| PC4 | -14.80 (-71.92, 42.31) | 0.598 |
| PC5 | -2.73 (-13.71, 8.26) | 0.613 |
| $E(M_{DMRmean})= \alpha+ \beta_1 ModeofDelivery+\varepsilon$ | | |
| | *Coefficient (95% Confidence Interval)* | *p-value* |
| Mode of delivery | -0.032 (-0.165, 0.100) | 0.619 |



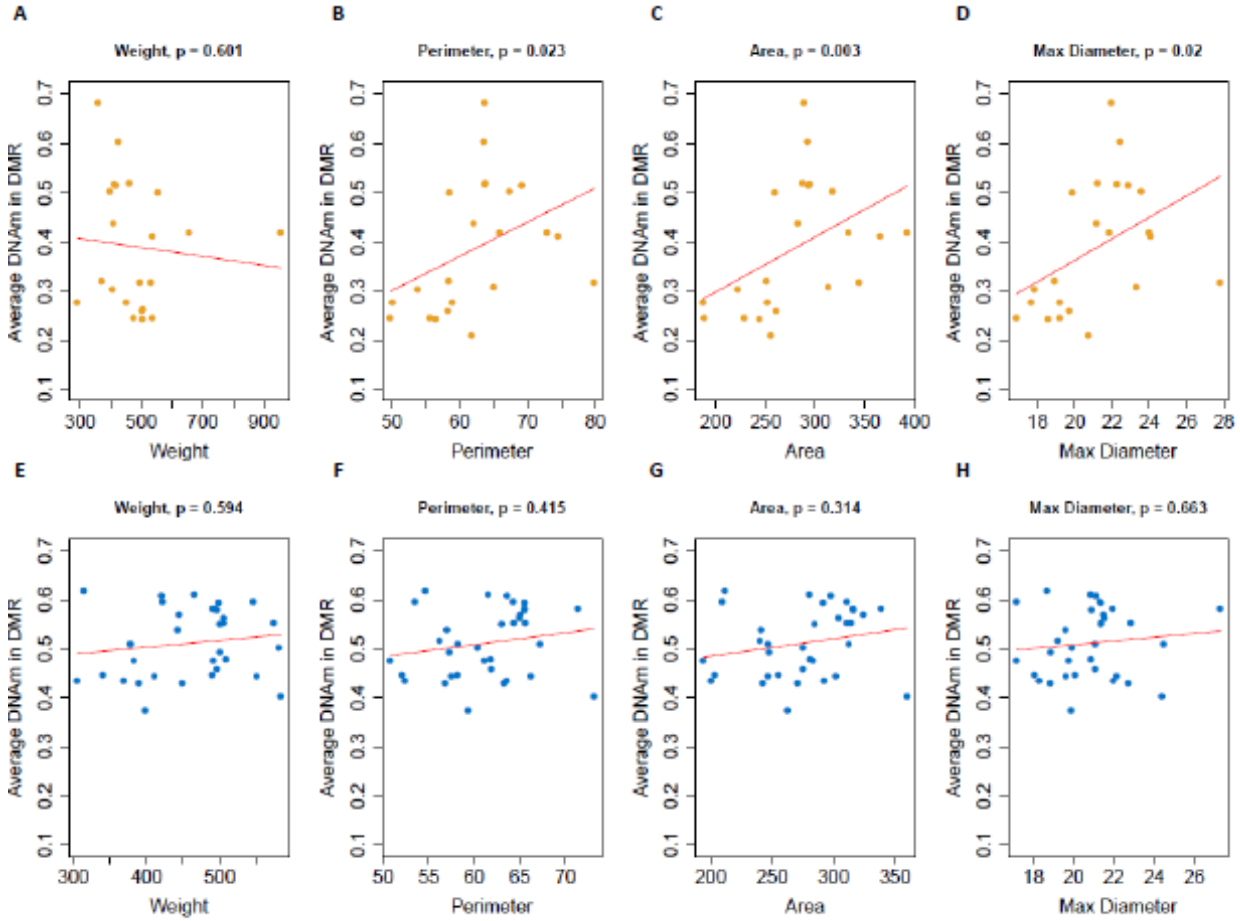

**Supplementary Figure V-1: Genome-wide results for DMR identification via BSmooth.** Y-axis depicts width (bp) of null (black dots) or candidate (green dots) DMRs. X-axis depicts area statistics (sum of t-statistics in DMR) of null or candidate DMRs. Null DMRs are those generated across 1000 permutations.



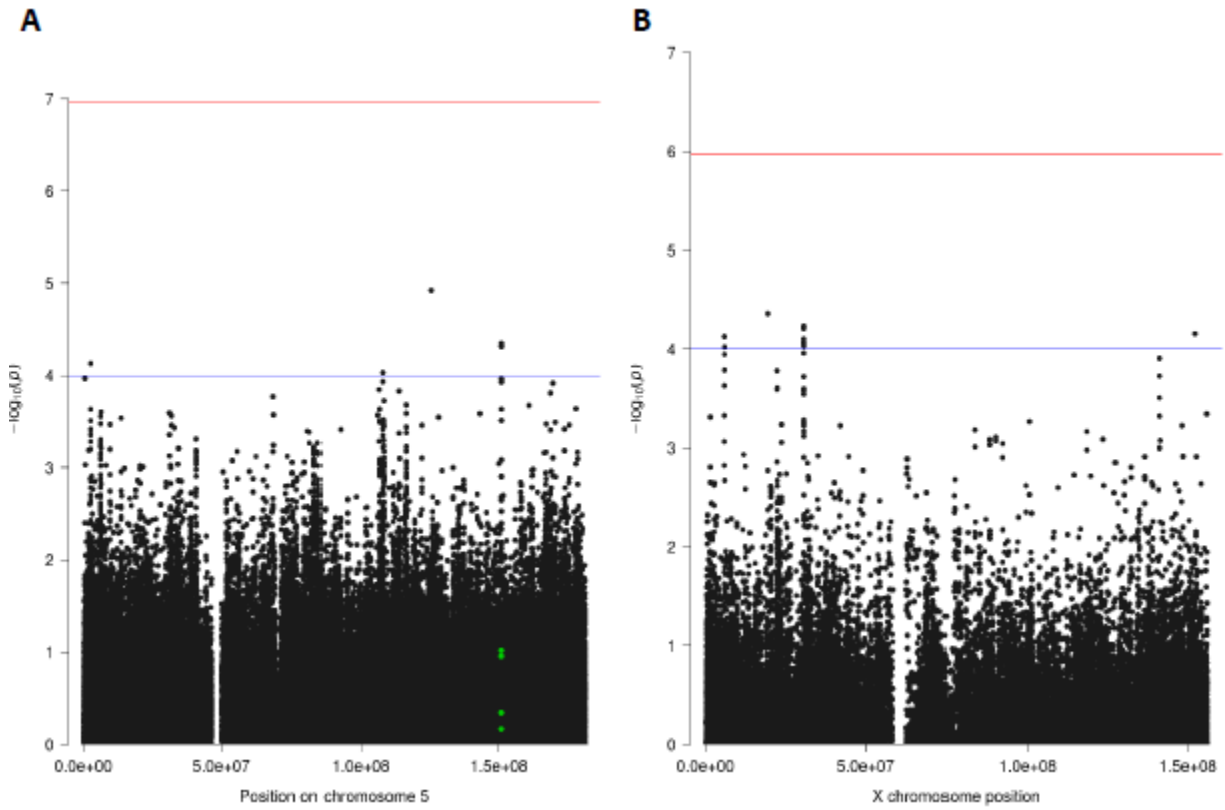

**Supplementary Figure V-2: Distribution of promoter-mapping TFBSs across the genome.** *ZNF300* count of 53 shown via red line, the 70$^{th}$ percentile of the distribution.



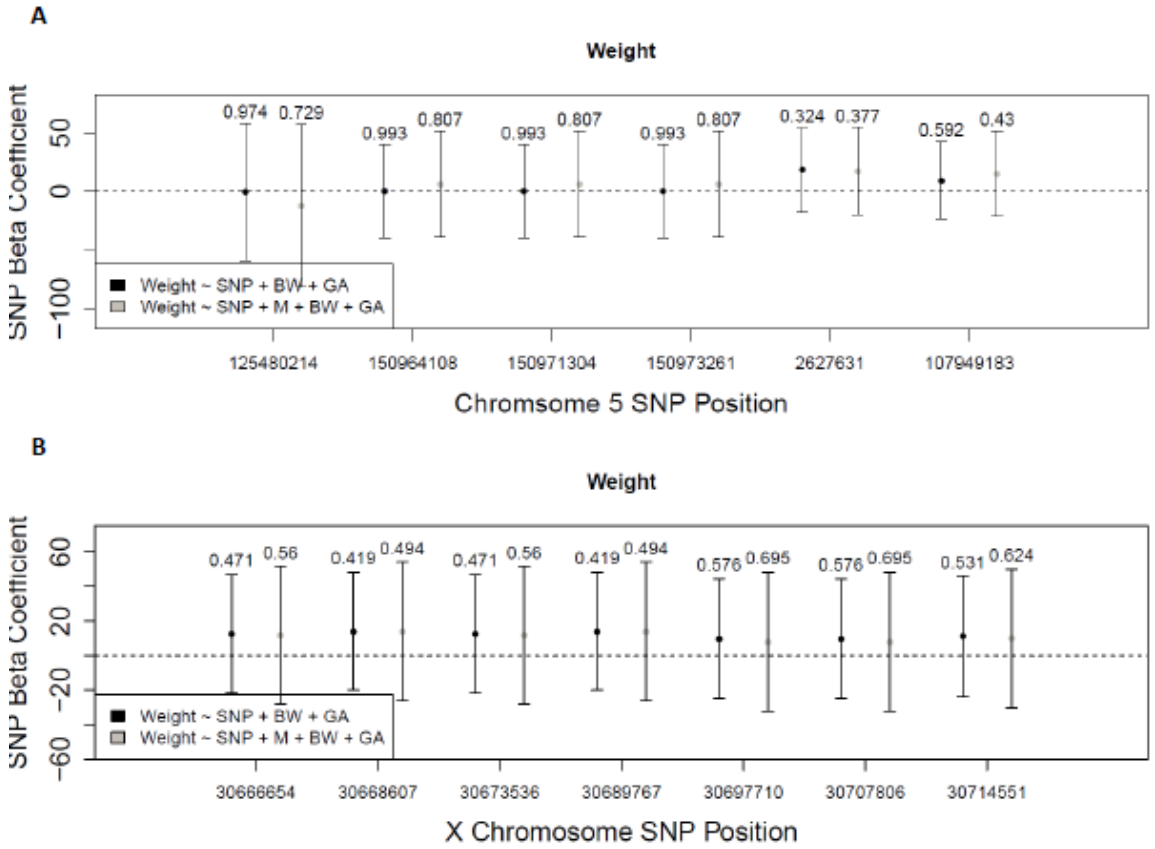

**Supplementary Figure V-3: Attempted replication of Roifman et al. finding relating *ZNF300* promoter methylation to IUGR.** **Panel A:** Percent methylation (y-axis) vs genomic position (x-axis) for WGBS EARLI samples (n = 74), samples colored by SGA (red, n = 6) and non-SGA (blue, n=68) status. **Panel B:** Same as A), but males only ($n_{SGA}$ = 3, $n_{total}$ = 34). **Panel C:** Same as A), but females only ($n_{SGA}$ = 3, $n_{total}$ = 34). **Panel D:** Percent methylation (y-axis) vs genomic position (x-axis) for 450K PR3 (GSE71678) samples (n = 338), samples colored by SGA (red, n = 19) and non-SGA (blue, n = 319) status. **Panel E:** Same as D), but males only ($n_{SGA}$ = 8, $n_{total}$ = 180). **Panel F:** Same as D), but females only ($n_{SGA}$ = 11, $n_{total}$ = 158).



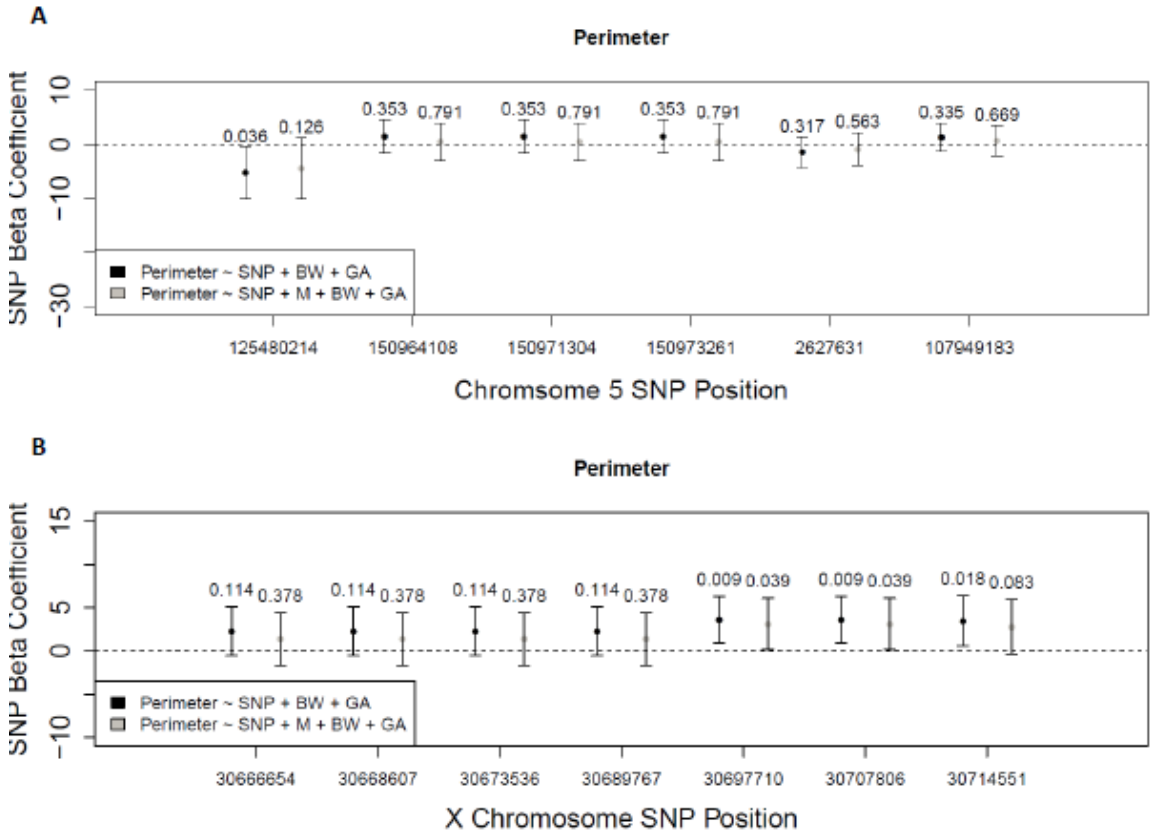

**Supplementary Figure V-4: Relationship between mean methylation levels at *ZNF300* DMR and placenta morphological features in males and females separately.** Mean DNAm levels in DMR (y-axis) vs. placental weight, perimeter, area, and maximum diameter (left to right, x-axis). P-values shown are for the regression of average DNAm in DMR onto each feature, adjusted for gestational age and birth weight. **A-D**) Females **E-G**) Males



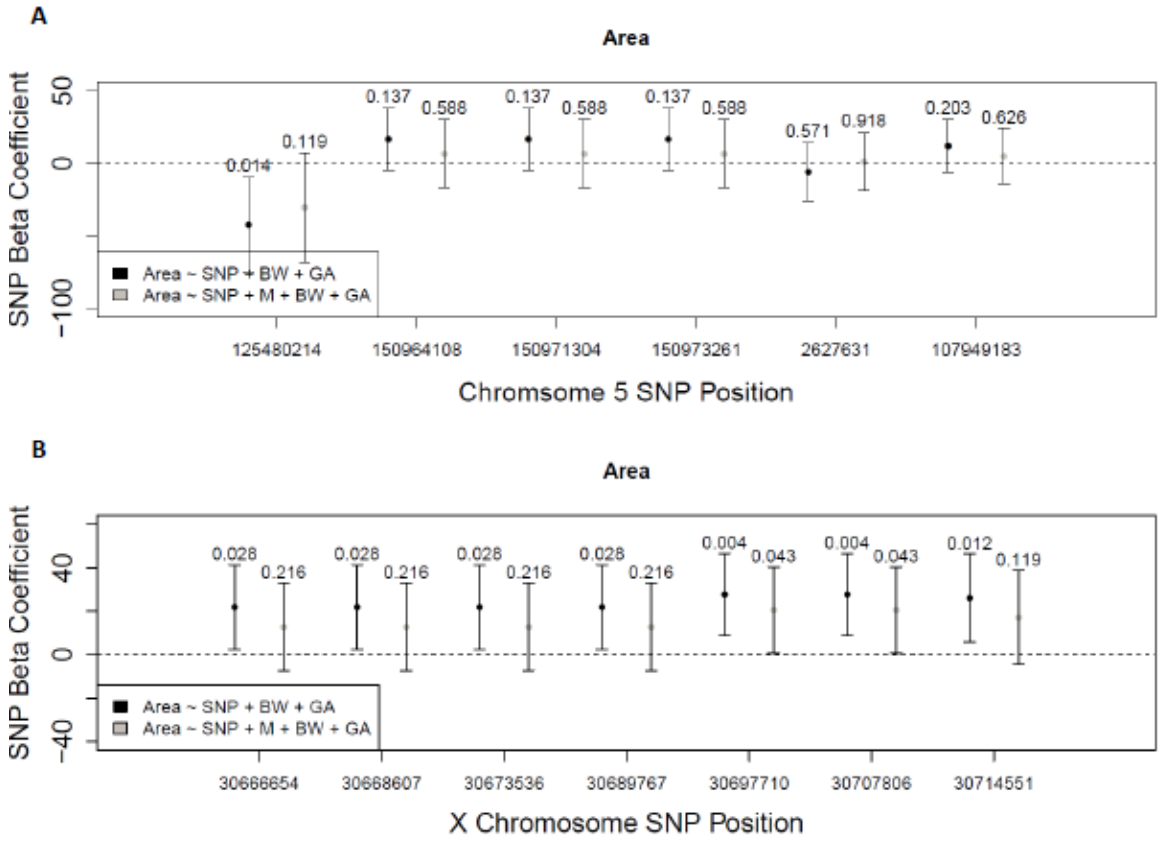

**Supplementary Figure V-5: Manhattan plots for methylation quantitative trait loci query for SNPs related to mean *ZNF300* DMR methylation levels. Panel A:** Chromosome 5 SNPs. **Panel B:** Chromosome X SNPs.



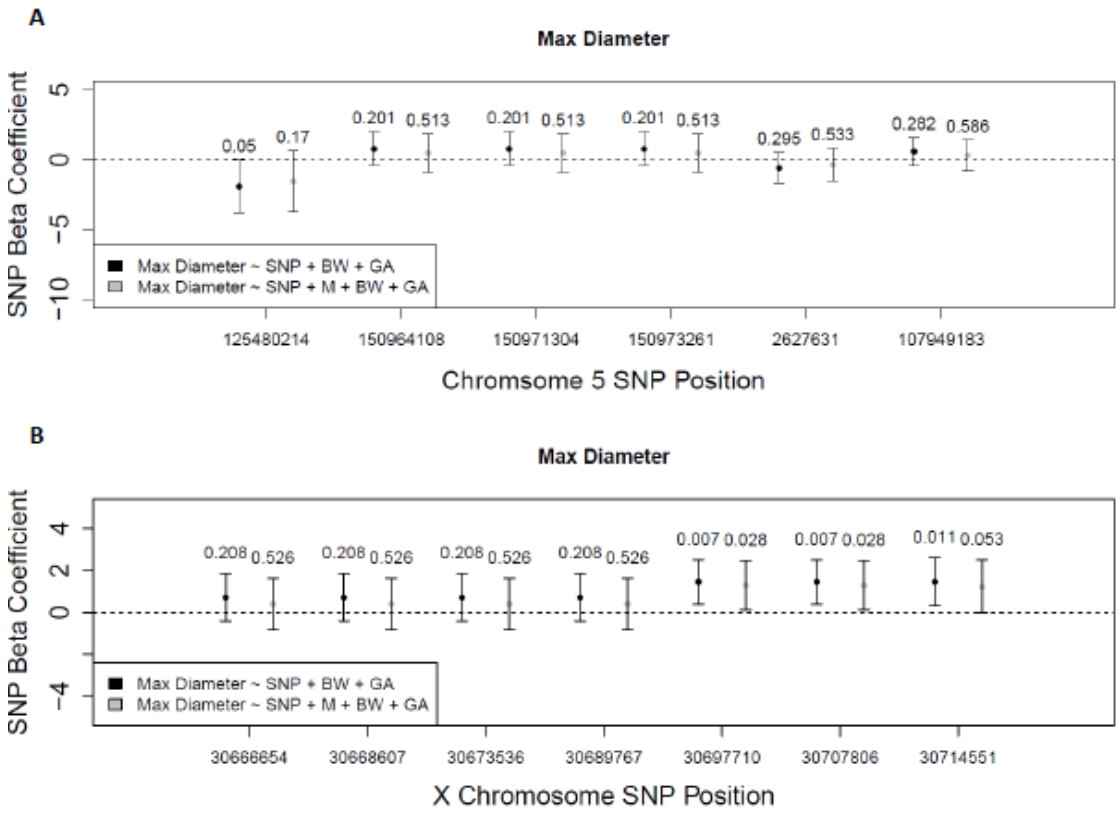

**Supplementary Figure V-6: Full mediation analysis (SNP → Methylation → Placenta weight) results for chromosome 5 and chromosome X SNPs suggestively associated with *ZNF300* methylation levels**. Points shown indicate regression coefficients for SNP terms in a model with the morphological phenotype as an outcome regressed onto the SNP, birthweight and gestational age without (black points) and with (gray points) including average methylation in the *ZNF300* DMR region. P-values and error bars are included along with the points. In each figure, **A)** Results for chromosome 5 and **B)** Results from chromosome X.




**Supplementary Figure V-7: Full mediation analysis (SNP → Methylation → Placenta perimeter) results for chromosome 5 and chromosome X SNPs suggestively associated with *ZNF300* methylation levels**. Points shown indicate regression coefficients for SNP terms in a model with the morphological phenotype as an outcome regressed onto the SNP, birthweight and gestational age without (black points) and with (gray points) including average methylation in the *ZNF300* DMR region. P-values and error bars are included along with the points. In each figure, **A)** Results for chromosome 5 and **B)** Results from chromosome X.




**Supplementary Figure V-8: Full mediation analysis (SNP → Methylation → Placenta area) results for chromosome 5 and chromosome X SNPs suggestively associated with *ZNF300* methylation levels**. Points shown indicate regression coefficients for SNP terms in a model with the morphological phenotype as an outcome regressed onto the SNP, birthweight and gestational age without (black points) and with (gray points) including average methylation in the *ZNF300* DMR region. P-values and error bars are included along with the points. In each figure, **A)** Results for chromosome 5 and **B)** Results from chromosome X.




**Supplementary Figure V-9: Full mediation analysis (SNP → Methylation → Placenta maximum diameter) results for chromosome 5 and chromosome X SNPs suggestively associated with *ZNF300* methylation levels**. Points shown indicate regression coefficients for SNP terms in a model with the morphological phenotype as an outcome regressed onto the SNP, birthweight and gestational age without (black points) and with (gray points) including average methylation in the *ZNF300* DMR region. P-values and error bars are included along with the points. In each figure, **A)** Results for chromosome 5 and **B)** Results from chromosome X.

# CURRICULUM VITAE

**Shan V. Andrews**
1819 Golden Gate Ave. Apt #16
San Francisco, CA 94115
Phone: 845-826-5924 Email: sandre17@jhu.edu

---

*Education*

**Johns Hopkins Bloomberg School of Public Health** *Expected April 2017*
PhD, Human Genetics and Genetic Epidemiology
Fellow: Maryland Genetics, Epidemiology, and Medicine Training Program
Dissertation: Characterizing the placenta methylome and its relationship to autism spectrum disorder
Advisors: Drs. Christine Ladd-Acosta and M. Daniele Fallin

**Johns Hopkins Bloomberg School of Public Health** *May 2014*
MHS, Human Genetics and Genetic Epidemiology
Thesis: Technical and biological sources of methodical variation in DNA methylation signal as measured by the Illumina HumanMethylation450 BeadChip
Advisors: Drs. Christine Ladd-Acosta and M. Daniele Fallin

**University of Delaware** *May 2012*
Bachelor's of Chemical Engineering, Minor in Biomedical Engineering

*Research Experience*

**Research Assistant/Master's Student/PhD Candidate**
Department of Epidemiology, JHSPH *Summer 2013-Present*

- Contributed to several Bioconductor R packages for epigenetic data analysis, including functions to:
  - Implement valid statistical significance testing in identification of differentially methylated regions
  - Estimate cell composition from mixed-cell samples measured via Illumina methylation arrays
  - House Illumina methylation data on sorted cord blood cells
- Integrated genotype and DNA methylation data on the same samples to characterize methylation distributions where samples cluster into distinct groups
  - Performed preprocessing and quality control of Illumina methylation array data
  - Successfully integrated publically accessible databases (dbSNP, Gene Expression Omnibus) to further enhance study findings
  - Work published in ***Epigenetics & Chromatin***

- Performed first genome-wide integration of genetic and epigenetic data in the context of autism spectrum disorder
    - Performed genome wide query for methylation quantitative trait loci (meQTLs) in multiple tissue types
    - Designed and performed genome-wide enrichment tests on SNP and CpG level using novel statistical approaches
    - Utilized publically available databases (Gene Expression Omnibus, 1000 Genomes, ENCODE, Roadmap Epigenomics Project, UCSC Genome Browser) to enhance or support study findings
    - Work published in ***Nature Communications***
- Currently working with largest dataset of whole genome bisulfite sequencing samples generated to date
    - Performed all steps of analysis pipeline, including: alignment using Bowtie2 & Bismark softwares, preprocessing, quality control, methylation status calling, and association analysis
    - Identified differentially methylated regions associated with fetal sex and autism spectrum disorder diagnosis
    - Performed first genome-wide interrogation of methylation quantitative trait loci in placenta tissue
- Mentored genetic epidemiology masters student in basic R programming and epigenetic data analysis

**Rotation Student/Research Assistant**
Institute of Genetic Medicine, Johns Hopkins School of Medicine *Jan 2016 – March 2016*

- Completed initial steps to use CRISPR-Cas9 technique to introduce single nucleotide variant for investigation of *APOE* gene in HepG2 cell line

*Skills*

- Extensive coursework and experience in epidemiology, biostatistics, and bioinformatics
- Extensive experience in quality control, processing, and analysis of a variety of omics-level data, including:
    - Array-based genotype data
    - Array-based DNA methylation data
    - Sequencing-based DNA methylation data
- Extensive experience in data integration, in particular genomic and epigenomic data, in both methodological and applied contexts
- Proficient in R, Perl, and Unix
- Proficient in genomics/bioinformatics-related software tools: Bioconductor R packages, PLINK, bedtools samtools, Bowtie2, Bismark, etc.
- Contributing to and maintaining group-based computing projects, including Bioconductor R packages

- Extensive experience operating on high performance computing cluster (batch jobs, parallel computing, etc.)
- Knowledge of psychiatric phenotypes, particularly autism spectrum disorder
- Extensive teaching experience of R programming language and epigenetic data analysis

*Teaching Experience*

**Principles of Genetic Epidemiology I**
Teaching Assistant *September-October 2013 and 2014*

**Principles of Genetic Epidemiology II**
Teaching Assistant *October-December 2013*

**Methodologic Challenges in Epidemiologic Research**
Teaching Assistant *March-May 2014-2016*

**Applications and Analysis of Epigenetic Data in Public Health Research** *June 2015-2016*
Lab Instructor

*Guest Lectures*

**Methodologic Challenges in Epidemiologic Research**
"An Introduction to R" March 2014

*Software* (https://github.com/sandrews5)

***FlowSorted.CordBlood.450k* (R package): Illumina 450k data on sorted cord blood cells**
Role: Author and Maintainer

***bumphunter* (R package): Tools for finding bumps in genomic data**
Role: Contributor

***minfi* (R package): Analyze Illumina Infinium DNA methylation arrays**
Role: Contributor

*Published Manuscripts*

1. **Andrews SV**, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Feinberg AP, Arking DE, Ladd-Acosta C*, Fallin MD*. *Cross-tissue*

*integration of genetic and epigenetic data offers insight into autism spectrum disorder*. Nature Communications, *8*.

2. **Andrews SV**†, Ladd-Acosta C†, Feinberg AP, Hansen KD, Fallin MD (2016). *"Gap hunting" to characterize clustered probe signals in Illumina methylation array data.* Epigenetics & Chromatin, *9*(1), 56.
3. Bakulski KM†, Feinberg JI†, **Andrews SV**, Yang J, Brown S, McKenney S, Witter F, Walston J, Feinberg AP, Fallin MD. *DNA methylation of cord blood cell types: applications for mixed cell birth studies*. Epigenetics, 11(5), 354-362.

*Published preprints*

1. **Andrews SV**, Sheppard B, Windham GC, Schieve LA, Schendel DE, Croen LA, Chopra P, Alisch RS, Newschaffer CJ, Warren ST, Feinberg AP, Fallin MD, Ladd-Acosta C. *Peripheral blood DNA methylation and autism spectrum disorder*. bioRxiv, 2018.
2. Hannon E, Schendel D, Ladd-Acosta C, Grove J, iPSYCH-Broad ASD Group, Soholm Hansen C, **Andrews SV**…et al. *Elevated polygenic burden for autism is associated with differential DNA methylation at birth*. bioRxiv, 2017.
3. **Andrews SV**, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Feinberg AP, Arking DE, Ladd-Acosta C*, Fallin MD*. *Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder*. bioRxiv, 2016. Published in Nature Communications, 2017.
4. **Andrews SV**†, Ladd-Acosta C†, Feinberg AP, Hansen KD, Fallin MD (2016). *"Gap hunting" to characterize clustered probe signals in Illumina methylation array data.* bioRxiv, 2016. Published in Epigenetics & Chromatin, 2016.

*Manuscripts in progress*

1. **Andrews SV**, Bakulski KM, Feinberg JI, Tryggvadottir R, Brown SC, Croen LA, Hertz-Picciotto I, Ladd-Acosta C, Hansen KD, Newschaffer CJ, Feinberg AP*, Fallin MD*. *Differential placenta methylation by fetal sex*. Manuscript in preparation.
2. **Andrews SV**, Bakulski KM, Feinberg JI, Tryggvadottir R, Brown SC, Croen LA, Hertz-Picciotto I, Ladd-Acosta C, Hansen KD, Newschaffer CJ, Feinberg AP*, Fallin MD*. *Genetic control of the placenta methylome*. Manuscript in preparation.

3. **Andrews SV**, Bakulski KM, Feinberg JI, Tryggvadottir R, Brown SC, Croen LA, Hertz-Picciotto I, Ladd-Acosta C, Hansen KD, Newschaffer CJ, Fallin MD*, Feinberg AP*. *Placenta methylation and autism spectrum disorder*. Manuscript in preparation.

*Presentations (presenting author indicated with *)*

1. **Andrews SV***, Ladd-Acosta C, Bakulski KM, Feinberg JI, Tryggvadottir R, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Yao R, Salafia CM, Feinberg AP, Hansen KD, Fallin MD. Placenta DNA methylation is associated with fetal sex at *ZNF300*. Presented at 2017 Epigenomics of Common Diseases Meeting (Poster Presentation & Lightning Talk).
2. Brucato M*, **Andrews SV**, Qian Y, Windham G, Schendel D, Schieve LA, Newschaffer C, Feinberg AP, Croen LA, Fallin MD, Ladd-Acosta C. Epigenetic alterations in childhood reflect prenatal exposure to maternal infection. Presented at 2017 Epigenomics of Common Diseases Meeting (Poster Presentation & Lightning Talk).
3. **Andrews SV***, Bakulski KM, Feinberg JI, Tryggvadottir R, Brown SC, Croen LA, Hertz-Picciotto I, Ladd-Acosta C, Hansen KD, Newschaffer CJ, Feinberg AP, Fallin MD. Placenta methylation and autism risk in the Early Autism Risk Longitudinal Investigation (EARLI). Presented at 2016 International Meeting for Autism Research. (Poster Presentation)
4. **Andrews SV***, Ellis SE, Bakulski KM, Feinberg AP, Arking DE, Ladd-Acosta C, Fallin MD. *Using methylation quantitative trait loci to enhance GWAS results for autism spectrum disorder across developmental stage and tissue type*. Presented at 2015 International Genetic Epidemiology Society Meeting. (Poster Presentation)
5. **Andrews SV***, Ellis SE, Bakulski KM, Feinberg AP, Arking DE, Ladd-Acosta C, Fallin MD. *Enrichment of methylation quantitative trait loci among genes associated with autism spectrum disorder*. Presented at 2015 International Meeting for Autism Research. (Oral Presentation)
6. Fallin MD*, **Andrews SV**, Lee B, Newschaffer C, Windham G, Schieve L, Croen L, Feinberg A, Ladd-Acosta C. *Assessment of sources of methylation variation and their*

*relationship to autism spectrum disorders*. Presented at 2014 International Meeting for Autism Research. (Poster Presentation)

7. Ladd-Acosta C*, Lee B, **Andrews SV**, Newschaffer C, Schieve L, Windham G, Croen L, Feinberg A, Fallin MD. *DNA methylation as a biomarker for prenatal exposures implicated in autism spectrum disorders*. Presented at 2014 International Meeting for Autism Research. (Poster Presentation)
8. **Andrews SV***, Croen L, Schieve L, Hansen K, Lee B, Newschaffer C, Feinberg A, Ladd-Acosta C, Fallin MD. *Blood-brain DNA methylation concordance in autism spectrum disorders*. Presented at 2014 International Meeting for Autism Research. (Poster Presentation)
9. **Andrews SV***, Ladd-Acosta C, Feinberg A, Fallin MD. *Identification of technical and biological sources of methodical variation in methylation signal as measured by the Illumina HumanMethylation450 BeadChip*. Presented at 2013 Epigenomics of Common Diseases Conference and 2013 Program in Quantitative Genomics Conference. (Poster Presentation)
10. Benke K*, **Andrews SV**, Fallin MD, Maher B. *Polygenic risk score associations may be improved with simple procedures*. Presented at 2013 International Genetic Epidemiology Society Conference. (Oral Presentation)

*Academic Service*

**Epidemiology Student Organization**
Masters Student Representative *September 2013-May 2014*

Doctoral Student Representative *September 2014-May 2015*

*Awards*

**1st Place (Master's students) – Genetics Research Day Poster Competition**
JHSPH and JHMI Institute of Genetic Medicine *February 2014*

**Charlotte Ferencz Fellowship**
Department of Epidemiology, JHSPH *March 2014*

**Honorable Mention (Doctoral students) – Genetics Research Day Poster Competition**
JHSPH and JHMI Institute of Genetic Medicine *February 2016*

**Louis I. and Thomas D. Dublin Award** *March 2016*
Departments of Epidemiology and Biostatistics, JHSPH